EUROPEAN RESPIRATORY journal
OFFICIAL SCIENTIFIC JOURNAL OF THE ERS

Pseudomonas aeruginosa adaptation and diversification in the non-Cystic Fibrosis bronchiectasis lung

Journal:	*European Respiratory Journal*
Manuscript ID	ERJ-02108-2016.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Hilliam, Yasmin; University of Liverpool, Institute of Infection and Global Health Moore, Matthew; University of Liverpool, Institute of Infection and Global Health Lamont, Iain; University of Otago, Department of Biochemistry Bilton, Diana; Royal Brompton Hospital and Imperial College, Haworth, Charles; Papworth Hospital, ; Foweraker, Juliet; Papworth Hospital, Department of Microbiology Walshaw, Martin; Liverpool Heart and Chest Hospital NHS Trust Williams, David; University of Liverpool, Institute of Integrative Biology Fothergill, Joanne; University of Liverpool, Institute of Infection & Global Health De Soyza, Anthony; University of Newcastle, ICM Winstanley, Craig; University of Liverpool, Institute of Infection and Global Health
Key Words:	bronchiectesis, pseudomonas, infections, genomics, evolution

REVISED MANUSCRIPT

Title: *Pseudomonas aeruginosa* adaptation and diversification in the non-Cystic Fibrosis bronchiectasis lung

Authors:

Yasmin Hilliam[1$], Matthew P. Moore[1$], Iain L. Lamont[2], Diana Bilton[3], Charles S. Haworth[4], Juliet Foweraker[4], Martin J. Walshaw[5], David Williams[1,6], Joanne L. Fothergill[1], Anthony De Soyza[7]*, Craig Winstanley[1]*

Affliliations:

1. Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

2. Department of Biochemistry, University of Otago, New Zealand

3. Royal Brompton Hospital, London, UK

4. Cambridge Centre for Lung Infection, Papworth Hospital, Cambridge, UK

5. Liverpool Heart and Chest Hospital, Liverpool, UK

6. Institute of Integrative Biology, University of Liverpool, Liverpool, UK

7. Institute for Cellular Medicine, Newcastle University, Newcastle-upon-Tyne, UK

8. Adult Bronchiectasis Service, Freeman Hospital, Newcastle-upon-Tyne, UK

*$ contributed equally

Corresponding author:

Professor Craig Winstanley

Institute of Infection and Global Health,

University of Liverpool,

Ronald Ross Building,

8 West Derby Street,

Liverpool L69 7BE, UK

Email: C.Winstanley@liv.ac.uk

Summary: In bronchiectasis, *P. aeruginosa* co-infections occur; bacterial populations both adapt and diversify by mutation

Abstract

To characterise *Pseudomonas aeruginosa* populations during chronic lung infections of non-Cystic Fibrosis bronchiectasis patients, we used whole genome sequencing to (i) assess the diversity of *P. aeruginosa* and the prevalence of multi-lineage infections, (ii) seek evidence for cross-infection or common source acquisition and (iii) characterize *P. aeruginosa* adaptations.

189 isolates, obtained from the sputa of 91 patients attending 16 adult UK bronchiectasis centres, were whole genome sequenced.

Bronchiectasis isolates were representative of the wider *P. aeruginosa* population. Of 24 patients where multiple isolates were examined, there were seven examples of multi-lineage infections, most likely arising from multiple infection events. The number of nucleotide variants between genomes of isolates from different patients was in some cases similar to the variations observed between isolates from individual patients implying the possible occurrence of cross infection or common source acquisition.

Our data indicate that during infections of bronchiectasis patients, *P. aeruginosa* populations adapt by accumulating loss of function mutations, leading to changes in phenotypes including different modes of iron acquisition and variations in biofilm-associated polysaccharides. The within-population diversification suggests that larger scale longitudinal surveillance studies will be required to capture cross infection or common source acquisition events at an early stage.

Abstract word count: 196

Introduction

Bronchiectasis is a chronic, progressive respiratory disease associated with irreversible widening of the bronchi [1]. Recent data suggest that in the UK incidence rates in women and men have risen to 35.2 and 26.9 respectively per 100,000 person-years [2]. In the USA the prevalence of adult bronchiectasis has been estimated at 52 in 100 000 people, with higher prevalence among women and older individuals [3]. Persistent *Pseudomonas aeruginosa* lung infections of bronchiectasis patients, occuring in approximately 30% of cases, are associated with poorer outcomes and premature mortality [4, 5].

The study of chronic *P. aeruginosa* lung infections has focused on cystic fibrosis (CF)-associated bronchiectasis, where patients are diagnosed, monitored and subjected to antibiotic therapy from a very early age. This contrasts with non-CF bronchiectasis patients, who present at a much older age and often have a shorter history of therapeutic interventions. Hence, bacterial isolates from non-CF bronchiectasis patients exhibit less resistance to antibiotics compared to isolates from adult CF patients [6]. Previous studies have characterized the evolution of *P. aeruginosa* during chronic lung infections in CF patients [7, 8]. More recently, high-resolution analyses have revealed extensive heterogeneity within *P. aeruginosa* populations in the CF lung [9-12], including the co-existence of multiple divergent lineages [13].

In CF, a number of transmissible strains of *P. aeruginosa* have been identified, leading to the introduction of measures to control cross infection [14]. The study of *P. aeruginosa* in relation to non-CF bronchiectasis is less advanced. In our single centre study of 50 *P. aeruginosa* isolates from 40 bronchiectasis patients using molecular typing, there was no compelling evidence for cross infection or a dominant clone [15]. However, whole genome sequence analysis of multiple bronchiectasis isolates has not been carried out. Here, we report the use of genomics to assess the diversity of *P. aeruginosa* strains causing

infections in non-CF bronchiectasis across multiple UK centres, to identify multi-strain infections, and to look for evidence for cross-infection or common source acquisition. We also characterise adaptive mutations and present evidence for within-population divergence during *P. aeruginosa* chronic lung infections of bronchiectasis patients.

Methods

Patients and bacterial isolates

The 189 *P. aeruginosa* isolates used in this study (see Table S1 in the online supplementary material) were isolated from sputum samples obtained from 93 patients with bronchiectasis and chronic *P. aeruginosa* infection (defined as two or more positive respiratory tract cultures in the preceding 12 months) attending 16 adult bronchiectasis centres throughout England and Wales. These included isolates collected as part of a multi-centre nebulized antibiotic trial [16], where patients were enrolled within 21 days of completing a course of antipseudomonal antibiotics for an exacerbation. Additional isolates from Newcastle (n = 8) and Liverpool (n = 53) were collected during observational studies. The methodology used for isolating *P. aeruginosa* from patient sputum samples is described in the online supplementary material.

For 24 patients, sets of isolates (two or more) from the same sample were analysed to look for evidence of multi-lineage infections. For three of these patients (patients 147 – 149), sets of 14 or 15 isolates from a single sample were sequenced for higher resolution analysis of within-population heterogeneity. For some analyses, to avoid biases arising from inclusion of multiple clonal genomes from the same patient, a subset of 99 genomes from 91 patients was used. This subset consisted of one randomly selected genome per clonal lineage per patient (see Table S1 in the online supplementary material). We use the term "clonal lineage" to describe isolates with shared multilocus sequence type (MLST) profile and

clustering according to core genome single nucleotide polymorphism (SNP)-based phylogeny.

DNA preparation and whole genome sequencing

Details of the extraction of genomic DNA from *P. aeruginosa* isolates, library preparation and whole-genome shotgun sequencing using Illumina short read sequencing technology are given in the online supplementary material. The European Nucleotide Archive accession number for the study is PRJEB14952.

Methods used for genome sequence assembly, extraction of MLST data, phylogenetic reconstruction using the core genome, and variant calling by mapping to the genome of PAO1 [17] to identify single nucleotide polymorphism (SNP) or small insertions or deletions (INDELs) are described in online supplementary material.

Identification of large deletions and virulence factor genes

Genome sequences were aligned to the reference genomes *P. aeruginosa* PAO1 (NC_002516, [17]) and *P. aeruginosa* LESB58 (FM209186; [18]) and large clone-specific deletions (10 kb and above) were identified using BRIG [19]. The boundaries of deletions were determined by aligning the genome sequences with the *P. aeruginosa* PAO1 genome using Mauve [20], implemented as part of the Geneious package (www.geneious.com). The presence and absence of virulence factor genes in genome assemblies was determined using Blastable (https://www.github.com/bawee/blastable). The *Pseudomonas* genome database (beta.pseudomonas.com) [21] was used to facilitate analysis of gene function.

Results

Diversity of *P. aeruginosa* Non-CF BE isolates and evidence for *P. aeruginosa* multi-lineage co-infections

Core genome SNP phylogenetic analysis alongside a collection of 331 *P. aeruginosa* isolate genomes from diverse clinical sources [22], indicated that the bronchiectasis isolates were widely distributed (see Figure S1 in the online supplementary material). From the 189 isolates, it was possible to extract complete MLST profiles for 160 (see Tables S1 and S2 in the online supplementary material), with the most widespread sequence types (STs) being ST-253 (PA14-like [23], 14 patients, 8 centres), ST-179 (7 patients, 4 centres), ST-17 (Clone C [23], 5 patients, 3 centres), ST-252 (4 patients, 4 centres) and ST-260 (4 patients, 3 centres). Using core genome SNP phylogeny, previous studies have sub-divided the wider *P. aeruginosa* population into two major groups (group I, which includes strain PAO1, and group II, which includes strain PA14) and one minor group of mostly unrelated clonal lineages [24, 25]. Of a subset of 99 genomes consisting of one randomly selected genome per clonal lineage per patient, 71 were located in group I and 27 in group II (Figure 1). Based on a combination of MLST genotype and core genome SNP phylogeny, of the 24 patients from whose samples multiple isolates were examined, there were seven examples of multi-lineage infections. In one patient (patient 92), three distinct clonal lineages of *P. aeruginosa* were identified. In patients 42, 72, 73, 84, 85 and 148 there were two co-existing lineages (Figure 1).

Evidence for shared lineages causing infections in different patients attending the same centre

The core genome SNP phylogeny identified a number of examples where closely-related clonal lineages were isolated from more than one patient attending the same centre (see Table S3 in the online supplementary material). In order to obtain a higher resolution comparison,

these isolates were analysed using pairwise comparisons across their entire genomes (Table S3), identifying five instances where the genomes of isolates from different patients attending the same centre varied at fewer than 200 sites (C6/C7, C29/C30, C105/109, C139/C141, C156/C159; Figure 2). This level of genome similarity is greater than in some pairwise comparisons of contemporary isolates of the same lineage from the same sputum sample (see Table S3 in the online supplementary material; from 184 variant sites [C110/C111] to >750 variant sites [C125/C126]).

The draft genome sequences of the sub-set of 99 bronchiectasis isolates were examined for the presence of large (> 10kb) deletions. A total of 36 different deletions (25 over 100 kb), ranging in size from 11 to 300 kb and representing independent genetic events, were identified (see Table S4 in the online supplementary material). These were distributed across 28 genomes in the 99-member genome subset. Most genomes had only one deletion, although two (C54 and C164) had three deletions and four (A119, C4, C85 and C119) had two. In most cases, isolates of the same clonal lineages from the same patient shared the same deletions. However, in patients 45, 55, 79 and 92 not all isolates of the same lineage had the same deletion. The genomes of isolate pairs C6/C7, C29/C30, C105/109, C139/C141 and C156/C159, which are from different patients but vary at fewer than 200 sites (Table 1), were indistinguishable by BRIG analysis (example shown in Fig. 3B).

Genomic diversity of isolates within patients can be similar to diversity between patients

In order to further assess the within-patient diversification exhibited by *P. aeruginosa* populations, larger sets of isolates from single sputum samples were analysed for three patients : 147 (15 isolates), 148 (15 isolates) and 149 (14 isolates) (Table 1). For two of these patients, the *P. aeruginosa* population was comprised of a single clonal lineage. For patient 148, two distinct clonal lineages were identified and these two sets of isolates were analysed

separately. In all four isolate sets analysed, the maximum pairwise SNP variations between two isolates of the same lineage exceeded 300, with a median of 179 or greater (Table 1), indicating the occurrence of within-patient diversification.

Loss of function mutations and deletions identified in multiple isolates

We used variant calling approaches to identify independent occurrences of loss of function mutations amongst the sub-set of 99 bronchiectasis isolate genomes. This yielded a number of examples of genes with known functions carrying independent loss of function mutations in multiple isolates (Table 2; see Table S5 in the online supplementary material). These include genes linked to mucoidy, virulence, osmoprotection, biofilm formation, motility, DNA repair and antimicrobial resistance (Table 2). The genes encoding all three components of the MexAB-OprM efflux pump appear amongst the most common loss of function mutations. Multiple isolates also carried loss of function mutations in genes encoding regulators (including *lasR*, *algU*, *fleR*, *vfr*). Among the 99 bronchiectasis isolates, the number of genes with loss of function mutations as listed in Table 2 ranged from 0 to 6 (see Figure S2 and Table S6 in the online supplementary material).

Hypermutability is a common trait amongst CF isolates of *P. aeruginosa*. Of the 99 panel isolates, 11 carried loss of function mutations in the DNA mismatch repair genes, *mutS* or *mutL* (see Table S1 in the online supplementary material). All but two of these were confirmed as having the hypermutable phenotype.

An alignment of all of the genomes containing deletions >10 kb relative to the genome of strain PAO1 revealed a strikingly non-random distribution, with 30 of the 36 deletions lying within the 1.9 to 2.8 Mb portion of the strain PAO1 genome. Genes within

this region include the *psl* genes, encoding an extracellular polysaccharide [26], genes encoding the siderophore pyoverdine, and genes encoding a type VI secretion apparatus [27].

We next specifically examined one representative of each of the 99 clonal lineages for the presence or absence of genes associated with pathogenicity (see Table S6 in the online supplementary material). 23 of these genomes lacked one or more of the *psl* genes. In contrast, all of the genomes contained all of the *alg* genes required for making alginate and the *pel* genes required for making Pel exopolysaccharide. Eleven of the genomes lacked genes required for synthesis of pyoverdine, with nine of these also lacking an *fpvA* receptor gene for uptake of ferripyoverdine, although the genes required for synthesis of an alternative siderophore, pyochelin, were present in all cases. Eleven of the genomes also lacked two or more genes of the Type VI secretion system [PA2360 (*hsiA3*) – PA2373 (*vgrG3*)] (see Table S6 in the online supplementary material). These findings are consistent with the occurrence of deletions of the region of the genome containing Psl, pyoverdine and type VI secretion genes in multiple isolates, although in some isolates smaller deletions (< 10kb) were detected.

Discussion

We used whole genome sequencing to obtain a cross section of the diversity of *P. aeruginosa* strains causing infections in bronchiectasis in the UK. Our data suggest that the distribution of *P. aeruginosa* lineages found amongst the bronchiectasis isolate collection broadly represents what is present in the global *P. aeruginosa* population. In contrast to CF [14], we found no data to suggest that there is a widespread transmissible strain amongst the UK non-CF bronchiectasis community. However, our study did not include large numbers of patients from individual centres. Lineages such as PA14-like and Clone C, that are naturally more abundant in nature [23], were also amongst the most abundant in the bronchiectasis

collection. Because some lineages are naturally more abundant, their occurrence (based on MLST) in multiple patients is not necessarily indicative of common source or cross infection. Whole genome sequencing offers higher resolution than methods such as MLST, allowing us to address this issue.

In a previous comparison of paired isolates from patients within the same bronchiectasis centre, in most patients (9 of 10) the two isolates shared a common genotype, with one patient found to be infected with two strains simultaneously [15]. In this study, of 24 patients from whose samples multiple isolates were examined, seven had multi-lineage infections. Similar multi-lineage infections have also been reported in CF, generally associated with children [28]. A number of studies in CF have also demonstrated the phenotypic [9, 11, 12] and genomic [10, 13, 29] diversification of single lineage *P. aeruginosa* populations in the CF lung. Here, we show for the first time that similar diversification occurs during infections of non-CF bronchiectasis patients. Both the prevalence of multi-lineage infections, and the diversification that occurs during the infection process emphasise the need to be cautious when interpreting the analysis of sputum samples based on single isolates of *P. aeruginosa*.

We found several examples of isolates from patients attending the same centre, that not only shared the same clonal lineage, but also differed genomically by <200 sites. Genomic variations between isolates from the same patient sample revealed similar, and in some cases higher, levels of variation. The occurrence of isolates with very high genetic relatedness in different patients strongly implies that there has been common source acquisition or cross infection. The extent of the nucleotide variations differentiating two isolates will be dependent upon (i) the length of time since the transmission event and (ii) the rate of mutation of the *P. aeruginosa* population during the infection. Further studies will be

needed to better define the role of cross infection or common source acquisitions in this patient group.

There was clear evidence for bacterial adaptation to the lung environment by the accumulation of mutations and deletions, including loss of function mutations in genes identified previously as being commonly mutated in CF, such as *mucA* (mucoidy) and *lasR* (quorum sensing). It is worth noting, however, that mutations in genes encoding some of the regulators highlighted in previous CF studies (*mexT*, *retS*, *exsD*, *ampR*) were observed either infrequently (two *mexT* and two *ampR* mutants) or not at all (see Table S5 in the online supplementary material). Mutations in global regulators potentially affect numerous processes. In CF, the pathoadaptive genes identified in different studies have varied, suggesting that there are multiple routes to adaptation to the CF lung [7, 8], a scenario which is likely to apply also to non-CF bronchiectasis.

Loss of function mutations in genes encoding the MexAB-OprM efflux pump were common amongst the bronchiectasis isolates. Although generally thought of as a multidrug efflux system important for antibiotic resistance, this system has also been implicated in virulence [30]. Hence, although it may seem counterintuitive that *P. aeruginosa* should adapt by losing an antibiotic resistance-related efflux pump, it may be that the driver for selection is related to a function other than antibiotic efflux. In contrast, the loss of function mutations in *mexS* can be linked directly to antibiotic resistance, since mutations in *mexS* promote upregulation of the MexEF-OprN MDR efflux pump [31].

The prevalence amongst non-CF BE isolates of deletions in a specific genomic region encoding pyoverdine and Psl polysaccharide was higher than in a dataset of 331 *P. aeruginosa* clinical isolate genomes [22], where 22 genomes lacked one or more *psl* genes, only three lacked one or more of the pyoverdine synthesis genes, and only one did not have an *fpvA* receptor gene. *P. aeruginosa* can utilise multiple pathways for iron acquisition [32].

During chronic lung infections in CF, *P. aeruginosa* adapts by favouring the heme utilisation route for iron acquisition rather than the pyoverdine siderophore system [33]. Our observations suggest a similar adaptation in non-CF bronchiectasis.

In order to protect itself from hostile environmental conditions or host defences *P. aeruginosa* can produce three exopolysaccharides contributing to biofilm formation: alginate, Psl and Pel [26]. It has been suggested that Psl is a key surface attachment determinant [34], whereas in the CF lung free-floating biofilm structures may be more important [35]. Other mutations favouring the production of Pel rather than Psl include mutations in *bifA* [36], *rbdA* [37], *oprF* [38] and *ladS* [39]. Hence, overall our observations indicate that in non-CF BE chronic lung infections, the Pel and alginate exopolysaccharides are favoured over Psl.

Other common loss of function mutations (in *pilJ*, *chpA* and *fimV*) are implicated in lost or amended twitching motility, an adaptation also seen both in CF [8] and in an artificial sputum biofilm model [40], suggesting that this may be an adaptation related to the viscosity of the sputum environment.

Our study represents the first comparative genomics analysis of multiple *P. aeruginosa* isolates associated with chronic lung infections of non-CF bronchiectasis patients. Although a larger, more targeted study, analysing greater numbers of isolates per sample, would be needed to determine the true prevalence of multi-lineage infections, this observation does suggest that it is common for multiple *P. aeruginosa* lineages to co-exist in bronchiectasis infections. Our study also demonstrates that within-sample diversity can be comparable in scale to the genetic variations that occur between isolates from different patients attending the same centre. These observations suggest that there is an urgent need for more detailed and larger scale longitudinal studies in non-CF patients, and for surveillance that captures the diversity within centres and would identify cross infection or common

source acquisition events earlier, allowing measures to be taken in order to minimise the spread of this important pathogen.

Acknowledgements

This work was supported by the UK Medical Research Council as part of the BRONCH-UK grant (Reference MR/L011263/1) awarded to AdS as Principal Investigator, with CW amongst the Co-Investigators. IL would like to acknowledge funding support from Cystic Fibrosis New Zealand, the New Zealand Lotteries Board (Health) and the NHMRC (Australia). We would like to thank the investigators and patients involved in the PROMIS trial [16]. We also thank Paul Roberts (Royal Liverpool and Broadgreen University Hospitals NHS Trust) for technical assistance.

Figure legends.

Figure 1. Evidence for multi-lineage co-infections in seven patients. The figure shows a core genome SNP phylogeny for the sub-set of 99 isolates, confirming that all but one isolate (B113) clusters into one of two major groups. Each bronchiectasis centre is represented by a different colour. Arrows sharing the same colour indicate isolates that were obtained from the same patient. The three isolates from the same patient 92 sample are numbered 1-3.

Figure 2. Example pairwise comparisons between isolates sharing the same clonal lineage that were isolated from more than one patient attending the same centre. The number of SNP variations are indicated, with the number of IN-DEL variations shown in brackets. Full details are shown in Table S3 in the online supplementary material. The five examples where isolates shared fewer than 200 variant sites are highlighted in green. All

isolates of ST-244 from patients attending Centre 4 were compared, with similarity graded according to variant sites (<200, green; 200 – 3000 orange; >3000 red).

Figure 3. Examples of alignment of genomes of bronchiectasis strains with that of reference strain *P. aeruginosa* PAO1. Sequences identified as present (dark grey) or absent (white) in the genome of PAO1 are indicated. (A) Isolates of the same lineage (ST-253) from the same patient. From innermost to outermost, C95, C97, C98, C99, C96. A deletion present in isolate C96 only is highlighted (see arrow). (B) Pairs of isolates (from innermost to outermost C6 and C7; and C156 and C159) that both share the same clonal lineage but are from different patients attending the same hospital. Isolates C6 and C7 share a large deletion and isolates C156 and C159 share a smaller overlapping deletion (Supplementary Table S4), as indicated (see arrows). (C) Isolates of different lineages from the same patient. From innermost to outermost, A77, A80, A85 (all ST-175); A78, A81, A82 (all ST-17). A large deletion present in the ST-17 isolates is indicated by an arrow. The figures were generated using BRIG.

References

1. De Soyza A, Brown JS, Loebinger MR, Bronchiectasis R, Academic N. Research priorities in bronchiectasis. *Thorax* 2013: 68(7): 695-696.
2. Quint JK, Millett ER, Joshi M, Navaratnam V, Thomas SL, Hurst JR, Smeeth L, Brown JS. Changes in the incidence, prevalence and mortality of bronchiectasis in the UK from 2004 to 2013: a population-based cohort study. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 2016: 47(1): 186-193.
3. Weycker D, Edelsberg J, Oster G, Tino G. Association between risk of acute exacerbations and age in patients with bronchiectasis. *Chest* 2005: 128(4): 153s-153s.
4. Finch S, McDonnell MJ, Abo-Leyah H, Aliberti S, Chalmers JD. A Comprehensive Analysis of the Impact of Pseudomonas aeruginosa Colonization on Prognosis in Adult Bronchiectasis. *Ann Am Thorac Soc* 2015: 12(11): 1602-1611.
5. Chalmers JD, Aliberti S, Blasi F. Management of bronchiectasis in adults. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 2015: 45(5): 1446-1462.
6. McDonnell MJ, Jary HR, Perry A, MacFarlane JG, Hester KL, Small T, Molyneux C, Perry JD, Walton KE, De Soyza A. Non cystic fibrosis bronchiectasis: A longitudinal retrospective observational

cohort study of Pseudomonas persistence and resistance. *Respiratory medicine* 2015: 109(6): 716-726.

7. Marvig RL, Sommer LM, Jelsbak L, Molin S, Johansen HK. Evolutionary insight from whole-genome sequencing of Pseudomonas aeruginosa from cystic fibrosis patients. *Future microbiology* 2015: 10(4): 599-611.

8. Winstanley C, O'Brien S, Brockhurst MA. Pseudomonas aeruginosa Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. *Trends in microbiology* 2016: 24: 327-337.

9. Mowat E, Paterson S, Fothergill JL, Wright EA, Ledson MJ, Walshaw MJ, Brockhurst MA, Winstanley C. Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections. *AmJRespirCrit Care Med* 2011: 183(12): 1674-1679.

10. Jorth P, Staudinger BJ, Wu X, Hisert KB, Hayden H, Garudathri J, Harding CL, Radey MC, Rezayat A, Bautista G, Berrington WR, Goddard AF, Zheng C, Angermeyer A, Brittnacher MJ, Kitzman J, Shendure J, Fligner CL, Mittler J, Aitken ML, Manoil C, Bruce JE, Yahr TL, Singh PK. Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. *Cell Host Microbe* 2015: 18(3): 307-319.

11. Darch SE, McNally A, Harrison F, Corander J, Barr HL, Paszkiewicz K, Holden S, Fogarty A, Crusz SA, Diggle SP. Recombination is a key driver of genomic and phenotypic diversity in a Pseudomonas aeruginosa population during cystic fibrosis infection. *Sci Rep* 2015: 5: 7649.

12. Workentine ML, Sibley CD, Glezerson B, Purighalla S, Norgaard-Gron JC, Parkins MD, Rabin HR, Surette MG. Phenotypic heterogeneity of Pseudomonas aeruginosa populations in a cystic fibrosis patient. *PloS one* 2013: 8(4): e60225.

13. Williams D, Evans B, Haldenby S, Walshaw MJ, Brockhurst MA, Winstanley C, Paterson S. Divergent, coexisting Pseudomonas aeruginosa lineages in chronic cystic fibrosis lung infections. *American journal of respiratory and critical care medicine* 2015: 191(7): 775-785.

14. Fothergill JL, Walshaw MJ, Winstanley C. Transmissible strains of Pseudomonas aeruginosa in Cystic Fibrosis lung infections. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 2012.

15. De Soyza A, Perry A, Hall AJ, Sunny SS, Walton KE, Mustafa N, Turton J, Kenna DT, Winstanley C. Molecular epidemiological analysis suggests cross-infection with Pseudomonas aeruginosa is rare in non-cystic fibrosis bronchiectasis. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 2014: 43(3): 900-903.

16. Haworth CS, Foweraker JE, Wilkinson P, Kenyon RF, Bilton D. Inhaled colistin in patients with bronchiectasis and chronic Pseudomonas aeruginosa infection. *American journal of respiratory and critical care medicine* 2014: 189(8): 975-982.

17. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV. Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen. *Nature* 2000: 406(6799): 959-964.

18. Winstanley C, Langille MG, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, Sanschagrin F, Thomson NR, Winsor GL, Quail MA, Lennard N, Bignell A, Clarke L, Seeger K, Saunders D, Harris D, Parkhill J, Hancock RE, Brinkman FS, Levesque RC. Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of Pseudomonas aeruginosa. *Genome Res* 2009: 19(1): 12-23.

19. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics* 2011: 12: 402.

20. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004: 14(7): 1394-1403.

21. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. *Nucleic acids research* 2016: 44(D1): D646-653.
22. Kos VN, Deraspe M, McLaughlin RE, Whiteaker JD, Roy PH, Alm RA, Corbeil J, Gardner H. The resistome of Pseudomonas aeruginosa in relationship to phenotypic susceptibility. *Antimicrobial agents and chemotherapy* 2015: 59(1): 427-436.
23. Fischer S, Klockgether J, Moran Losada P, Chouvarine P, Cramer N, Davenport CF, Dethlefsen S, Dorda M, Goesmann A, Hilker R, Mielke S, Schonfelder T, Suerbaum S, Turk O, Woltemate S, Wiehlmann L, Tummler B. Intraclonal genome diversity of the major Pseudomonas aeruginosa clones C and PA14. *Environ Microbiol Rep* 2016: 8(2): 227-234.
24. Stewart L, Ford A, Sangal V, Jeukens J, Boyle B, Kukavica-Ibrulj I, Caim S, Crossman L, Hoskisson PA, Levesque R, Tucker NP. Draft genomes of 12 host-adapted and environmental isolates of Pseudomonas aeruginosa and their positions in the core genome phylogeny. *Pathog Dis* 2014: 71(1): 20-25.
25. Freschi L, Jeukens J, Kukavica-Ibrulj I, Boyle B, Dupont MJ, Laroche J, Larose S, Maaroufi H, Fothergill JL, Moore M, Winsor GL, Aaron SD, Barbeau J, Bell SC, Burns JL, Camara M, Cantin A, Charette SJ, Dewar K, Deziel E, Grimwood K, Hancock RE, Harrison JJ, Heeb S, Jelsbak L, Jia B, Kenna DT, Kidd TJ, Klockgether J, Lam JS, Lamont IL, Lewenza S, Loman N, Malouin F, Manos J, McArthur AG, McKeown J, Milot J, Naghra H, Nguyen D, Pereira SK, Perron GG, Pirnay JP, Rainey PB, Rousseau S, Santos PM, Stephenson A, Taylor V, Turton JF, Waglechner N, Williams P, Thrane SW, Wright GD, Brinkman FS, Tucker NP, Tummler B, Winstanley C, Levesque RC. Clinical utilization of genomics data produced by the international Pseudomonas aeruginosa consortium. *Frontiers in microbiology* 2015: 6: 1036.
26. Franklin MJ, Nivens DE, Weadge JT, Howell PL. Biosynthesis of the Pseudomonas aeruginosa Extracellular Polysaccharides, Alginate, Pel, and Psl. *Frontiers in microbiology* 2011: 2: 167.
27. Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, Goodman AL, Joachimiak G, Ordonez CL, Lory S, Walz T, Joachimiak A, Mekalanos JJ. A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus. *Science* 2006: 312(5779): 1526-1530.
28. Hall AJ, Fothergill JL, McNamara PS, Southern KW, Winstanley C. Turnover of strains and intraclonal variation amongst Pseudomonas aeruginosa isolates from paediatric CF patients. *Diagnostic microbiology and infectious disease* 2014: 80(4): 324-326.
29. Chung JC, Becq J, Fraser L, Schulz-Trieglaff O, Bond NJ, Foweraker J, Bruce KD, Smith GP, Welch M. Genomic variation among contemporary Pseudomonas aeruginosa isolates from chronically infected cystic fibrosis patients. *Journal of bacteriology* 2012: 194(18): 4857-4866.
30. Hirakata Y, Srikumar R, Poole K, Gotoh N, Suematsu T, Kohno S, Kamihira S, Hancock RE, Speert DP. Multidrug efflux systems play an important role in the invasiveness of Pseudomonas aeruginosa. *JExpMed* 2002: 196(1): 109-118.
31. Sobel ML, Neshat S, Poole K. Mutations in PA2491 (mexS) promote MexT-dependent mexEF-oprN expression and multidrug resistance in a clinical strain of Pseudomonas aeruginosa. *Journal of bacteriology* 2005: 187(4): 1246-1253.
32. Konings AF, Martin LW, Sharples KJ, Roddam LF, Latham R, Reid DW, Lamont IL. Pseudomonas aeruginosa uses multiple pathways to acquire iron during chronic infection in cystic fibrosis lungs. *Infection and immunity* 2013: 81(8): 2697-2704.
33. Marvig RL, Damkiaer S, Khademi SM, Markussen TM, Molin S, Jelsbak L. Within-host evolution of Pseudomonas aeruginosa reveals adaptation toward iron acquisition from hemoglobin. *mBio* 2014: 5(3): e00966-00914.
34. Colvin KM, Irie Y, Tart CS, Urbano R, Whitney JC, Ryder C, Howell PL, Wozniak DJ, Parsek MR. The Pel and Psl polysaccharides provide Pseudomonas aeruginosa structural redundancy within the biofilm matrix. *Environmental microbiology* 2012: 14(8): 1913-1928.

35. Bjarnsholt T, Jensen PO, Fiandaca MJ, Pedersen J, Hansen CR, Andersen CB, Pressler T, Givskov M, Hoiby N. Pseudomonas aeruginosa biofilms in the respiratory tract of cystic fibrosis patients. *PediatrPulmonol* 2009: 44(6): 547-558.

36. Kuchma SL, Brothers KM, Merritt JH, Liberati NT, Ausubel FM, O'Toole GA. BifA, a cyclic-Di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by Pseudomonas aeruginosa PA14. *Journal of bacteriology* 2007: 189(22): 8165-8178.

37. An S, Wu J, Zhang LH. Modulation of Pseudomonas aeruginosa biofilm dispersal by a cyclic-Di-GMP phosphodiesterase with a putative hypoxia-sensing domain. *Applied and environmental microbiology* 2010: 76(24): 8160-8173.

38. Bouffartigues E, Moscoso JA, Duchesne R, Rosay T, Fito-Boncompte L, Gicquel G, Maillot O, Benard M, Bazire A, Brenner-Weiss G, Lesouhaitier O, Lerouge P, Dufour A, Orange N, Feuilloley MG, Overhage J, Filloux A, Chevalier S. The absence of the Pseudomonas aeruginosa OprF protein leads to increased biofilm formation through variation in c-di-GMP level. *Frontiers in microbiology* 2015: 6: 630.

39. Records AR, Gross DC. Sensor kinases RetS and LadS regulate Pseudomonas syringae type VI secretion and virulence factors. *Journal of bacteriology* 2010: 192(14): 3584-3596.

40. Davies E, James CE, Williams D, O'Brien S, Fothergill JL, Haldenby S, Paterson S, Winstanley C, Brockhurst MA. Temperate phages both mediate and drive adaptative evolution in pathogen biofilms. *Proceedings of the National Academy of Sciences of the United States of America* 2016: in press.

Table 1. Summary of genomic diversity observed within the same clonal lineage of *P. aeruginosa* in individual patients

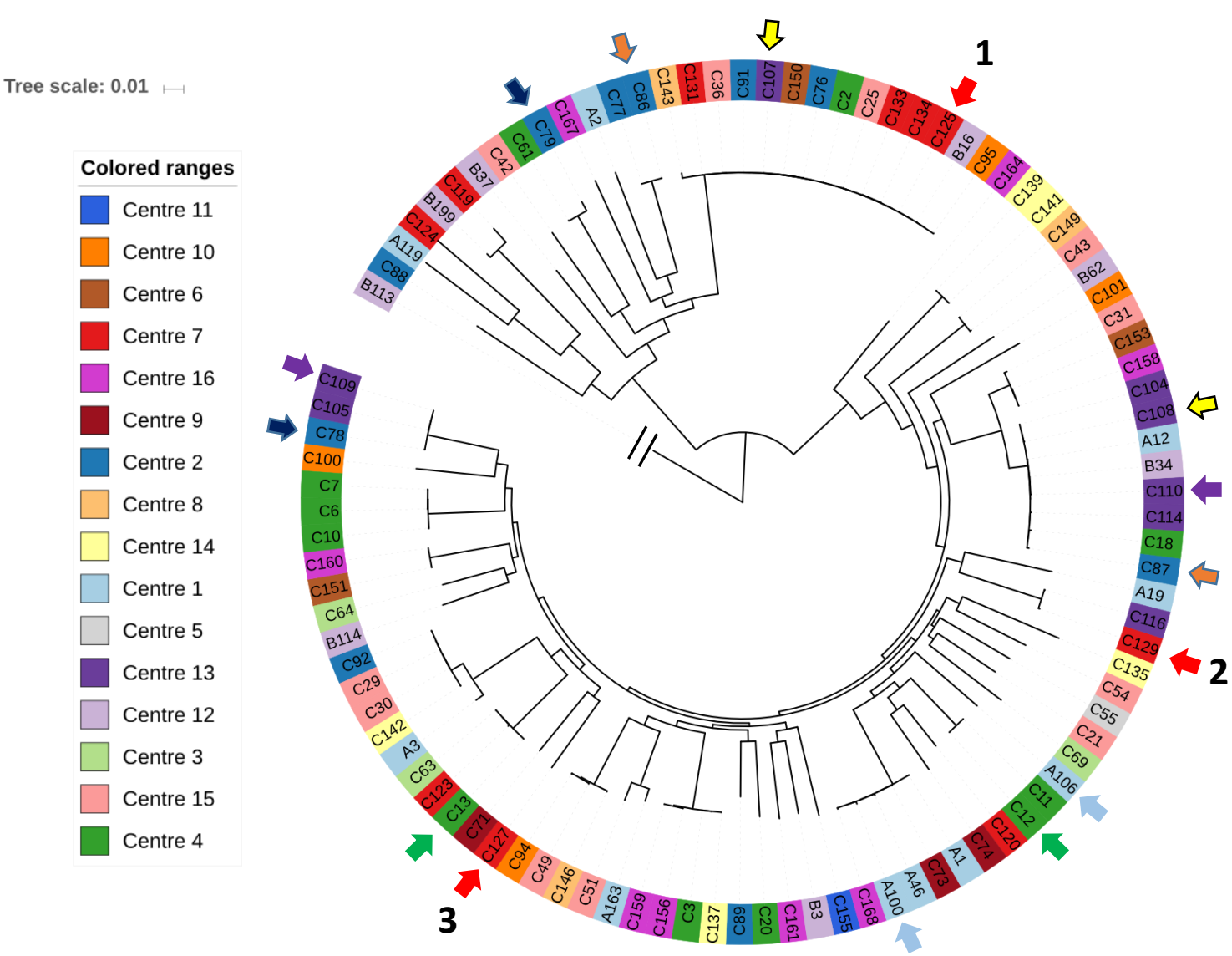
Patient	Number of isolates	Mean SNPs	Median SNPs	SNP range	Mean Indels	Median Indels	Indel range
Patient 147	15	336.35	261.00	88-640	15.20	14.00	0-35
Patient 148 (ST17)	4	451.50	482.50	159-654	23.83	25.50	6-34
Patient 148 (ST175)	11	195.45	179.00	79-403	9.27	5.00	0-36
Patient 149	14	209.01	206.00	68-327	11.40	10.00	3-28

The table indicates the number of single nucleotide polymorphisms (SNPs) and small insertion and deletion (INDEL) differences between the genomes of contemporary isolates from single sputum samples.

Table 2 Loss of function mutations occurring in multiple isolates. Only mutations predicted to lead to loss of function were included (ie. introduction of a stop codon, or a frame-shift mutation). The number of independent mutations indicates the number of isolates carrying unique mutations in the listed gene. The Table shows those genes where the number of independent occurrences of a mutation was equal to or greater than five.

Gene	PAO1 gene number	Number of independent occurrences of a mutation	Function / Comment
mexB	PA0426	16	Transporter from MexAB-OprM efflux pump, antibiotic resistance, virulence
mucA	PA0763	13	Anti-sigma factor, mutations can lead to mucoidy.
betT2	PA5291	9	Transporter, uptake of small molecules such as choline and glycine betaine, contributing to growth via phosphatidyl choline metabolism and osmoprotection
bifA	PA4367	7	Cyclic-di-GMP phosphodiesterase, inversely regulates biofilm formation
mexA	PA0425	7	Membrane fusion protein from MexAB-OprM efflux pump, antibiotic resistance, virulence
pcoA	PA2065	7	Copper resistance
PA4469	PA4469	7	Hypothetical protein encoded by a gene in same operon as and upstream of *sodM* (superoxide dismutase; response to oxidative stress)
rbdA	PA0861	7	Cyclic-di-GMP phosphodiesterase, modulation of biofilm dispersal, negative regulation of Pel production
pilJ	PA0411	6	Methyl accepting chemotaxis receptor-like protein involved in twitching motility and biofilm formation
oprM	PA0427	6	Outer membrane protein from MexAB-OprM efflux pump, antibiotic resistance, virulence
oprF	PA1777	6	Major porin, biofilm formation
chpA	PA0413	5	Chemotaxis-like chemosensory protein involved in twitching motility
fimV	PA3115	5	Peptidoglycan-binding protein, promotes type IV pilin assembly, twitching motility
ladS	PA3974	5	Sensor kinase, implicated in switch between acute and chronic infection
mutL	PA4946	5	Mismatch repair system, DNA repair, mutation can lead to mutator phenotype
gmd	PA5453	5	GDP-mannose 4,6-dehydratase,

mexS	PA2491	5	Mutations promote MexT-dependent *mexEF-oprN* expression and multidrug resistance
pchE	PA4226	5	Pyochelin synthesis
PA0054	PA0054	5	Hypothetical protein


Tree scale: 0.01
Colored ranges
Centre 11
Centre 10
Centre 6
Centre 7
Centre 16
Centre 9
Centre 2
Centre 8
Centre 14
Centre 1
Centre 5
Centre 13
Centre 12
Centre 3
Centre 15
Centre 4
1
2
3
C1 C107 C150 C76 C2 C25 C133 C134 C125 B16 C95 C164 C139 C141 C149 C43 B62 C101 C31 C153 C158 C104 C108 A12 B34 C110 C114 C18 C87 A19 C116 C129 C135 C54 C55 C21 C69 A106 C11 C12 C120 C74 A1 C73 A46 A100 C168 C155 B3 C161 C20 C89 C137 C3 C156 C159 A163 C51 C146 C49 C94 C127 C71 C13 C123 C63 A3 C142 C30 C29 C92 B114 C64 C151 C160 C10 C6 C7 C100 C78 C105 C109 B113 A119 C88 C124 B199 C119 B37 C42 C61 C79 C167 A2 C77 C86 C143 C131 C36

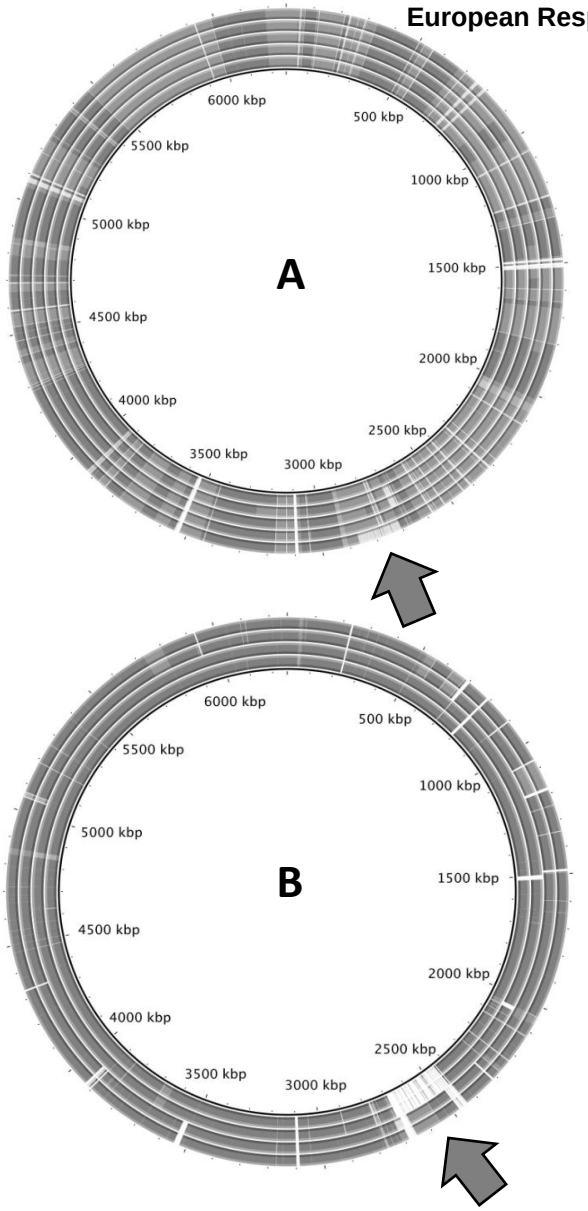
Centre 4 (ST-244)					
Patient:	38	39	40	40	40
Isolate	C6	C7	C8	C9	C10
C6		179(8)	3790(79)	3733(74)	3833(62)
C7			3736(82)	3714(75)	3929(65)
C8				281(5)	603(14)
C9					515(11)

Centre 15 (ST-252)		
Patient:	48	49
Isolate	C29	C30
C29		168(3)

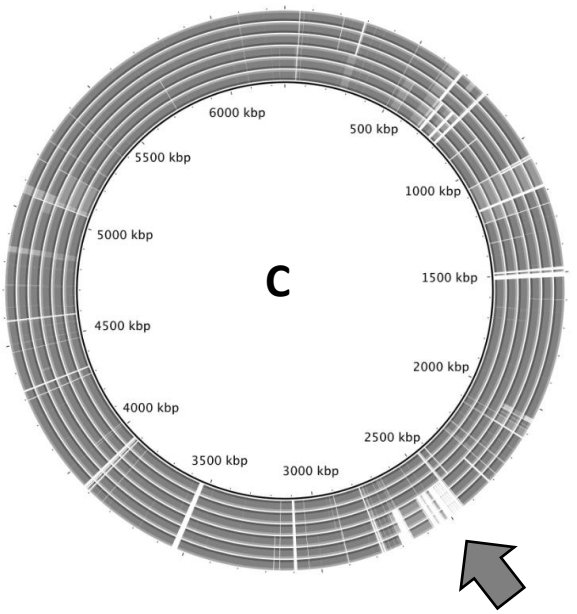
Centre 13 (ST-840)		
Patient:	83	85
Isolate	C105	C109
C105		131(8)

Centre 16 (ST-260)		
Patient:	108	110
Isolate	C156	C159
C156		160(3)

Centre 14 (ST-2102)		
Patient:	98	99
Isolate	C139	C141
C105		177(3)


A
6000 kbp
500 kbp
5500 kbp
1000 kbp
5000 kbp
1500 kbp
4500 kbp
2000 kbp
4000 kbp
2500 kbp
3500 kbp
3000 kbp

B
6000 kbp
500 kbp
5500 kbp
1000 kbp
5000 kbp
1500 kbp
4500 kbp
2000 kbp
4000 kbp
2500 kbp
3500 kbp
3000 kbp


C
6000 kbp
500 kbp
5500 kbp
1000 kbp
5000 kbp
1500 kbp
4500 kbp
2000 kbp
4000 kbp
2500 kbp
3500 kbp
3000 kbp

0.01

categories
Bronchiectasis isolates
Reference_strains

Online Supplementary Material for:

Pseudomonas aeruginosa adaptation and diversification in the non-Cystic Fibrosis bronchiectasis lung

Yasmin Hilliam, Matthew P. Moore, Iain L. Lamont, Diana Bilton, Charles S. Haworth, Juliet Foweraker, Martin J. Walshaw, David Williams, Joanne L. Fothergill, Anthony De Soyza, Craig Winstanley

Supplementary Methods

Isolation of *P. aeruginosa* from sputum samples

All samples were cultured for routine quantitative microbiology. Sputum was homogenised with equal parts of 0.1% (v/v) dithiothreitol, and the homogenised sample was diluted in sterile distilled water to 1 in 200 and 1 in 10,000. Both dilutions were spread by Whitley Automatic Spiral Plater (WASP) onto agar plates. Cultures on Columbia blood agar, chocolate agar and cysteine lactose electrolyte deficient agar (CLED) were incubated at 37^{o}C in 5% (v/v) CO_2 for up to 48 h. Cultures on Pseudomonas agar plus CFC supplement (PCFC) were incubated at 37^{o}C in air for up to 48 h and then checked for small colony variants (SCV) after incubating on the bench for up to a further 48 h. Colony forming unit (CFU) representatives of up to four colonial morphotypes were sub cultured and stored in 15% (v/v) Glycerol for archiving at -80^{o}C.

DNA preparation and genome sequencing

Genomic DNA was extracted from isolates using a Promega Wizard Genomic DNA Purification Kit, quantified using a Qubit 3.0 fluoromiter (Qubit dsDNA broad range assay kit, Life Technologies) and tested for purity using a NanoDrop 1000 spectrophotometer (Thermo Scientific). Library preparation and whole-genome shotgun sequencing was performed by the Centre for Genomic Research at the University of Liverpool, UK using Illumina short read sequencing technology. Shotgun libraries were prepared from the normalised samples using TruSeq Nano library preparation kit. Following library preparation, paired-end sequencing (2 x 100 bp) was performed by multiplexing into one lane of the Illumina HiSeq platform and sequenced with SBS V4 chemistry.

Following processing, the raw Fastq files were trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2.1 [1]. The option –O 3 was used so that the 3′ end of any reads which match the adapter sequence for 3 bp or more were trimmed. The reads were further trimmed using Sickle (https://github.com/najoshi/sickle) version 1.200 with a minimum window quality score of 20. Reads shorter than 10 bp after trimming were removed. If only one read of a pair passed this filter, it was included in the R0 file, with files R1 and R2 containing corresponding paired-end sequences. Quality filtered and adapter trimmed short reads were *de novo* assembled and scaffolded using the A5 MiSeq assembler [2]. Genome assembly quality metrics such as N50, largest contig, and overall number of contigs were produced using QUAST [3].

Core genome SNP phylogeny

The core genome was extracted using Panseq [4] and was defined as 500 bp fragments of all genomes in this study which matched with at least 85% similarity. A phylogenetic tree was approximated from core genome polymorphic sites, not including gaps or ambiguous bases by maximum likelihood with inner node bootstrap (n = 1000) and 10 discrete gamma

categories. All phylogenetic analyses were performed using MEGA6 [5] and visualised using the iTOL software [6]. Long branches were reduced for clarity.

Multilocus Sequence Typing

MLST profiles were extracted based on the pubMLST *Pseudomonas aeruginosa* scheme (http://pubmlst.org/paeruginosa/) using a specific tool (https://github.com/tseemann/mlst). It was not possible to extract complete MLST profiles from all genomes. In the context of this study, a lineage is defined on the basis of MLST profile and core genome SNP phylogeny.

Whole genome pairwise comparisons

Pairwise comparisons between assembled genomes were performed using MUMmer 3.0 [7]. Any positions in the alignment with ambiguous nucleotides were removed.

Read mapping and variant calling

All genome short read files (*fastq*) were mapped to reference genome PAO1 [8] using bwa-075a (mem) [9] producing a sequence alignment map (*sam*) files which were converted to binary alignment map files (*bam*) using picard tools-1.85 (https://broadinstitute.github.io/picard/). The reference genome was first indexed and sorted using bwa and SAMtools [10] respectively and a sequence dictionary created using picard tools-1.85. Variants were called following the Genome Analysis Toolkit (GATK) best practices workflow, as follows: duplicates were marked, the *bam* file indexed and sorted with picard tools-1.85, realignment targets created, INDELs realigned with GATK-3.3 [11] and variants called with the HaplotypeCaller module. Variants were filtered using vcffilter (https://github.com/vcflib/vcflib) with the standard parameters (DP >9, QUAL >10).

Resulting variant call files (*vcf*) were annotated to predict functional outcomes of variants compared with PAO1 genes using SnpEff [12]. Using SAMtools depth any variants within genes to which short reads had not aligned to 100% of its length were excluded from further analysis, leaving substitutions and short insertions or deletions (INDELs) and eliminating genes from functional variant analysis to which reads did not fully align due to sequencing error (lack of coverage) or genuine large deletions/complete absence.

Loss of function mutations

Predicted loss of function mutations were inferred to have been acquired independently in the population if they differed by position or type between genomes. Where the position and type were the same they were inferred to be shared; in genes where a mutation was shared with other genomes, further loss of function mutations were assumed to have been acquired since the common, ancestral acquisition of the shared mutation.

Supplementary Tables:

Supplementary Table S1. Bacterial isolates used in this study. ST refers to the designated multilocus sequence type. *novel shared MLST and they cluster according to SNP phylogeny. Isolates C21-C23 cluster together according to core genome SNP phylogeny. Isolates included in the subset of 99 genomes are highlighted in red. The Mutator column indicates the presence of mutations associated with hypermutability (INDEL, causing a frameshift or STOP, introduction of a stop codon). Of the 11 isolates carrying such a mutation, all but two (B113 and C78) were confirmed as having a hypermutable phenotype using assays reported previously [13]. Isolates labelled with the same letter from [a] to [i] were considered to be from the same lineage as each other on the basis of incomplete MLST profiles (see Table S2) and clustering by core genome SNP phylogeny (see Figure S1).

Isolate ID	Center	Date	Patient	ST	Mutator
A1	1	09/10/2014	1	17	
A2	1	16/10/2014	2	207	
A3	1	10/10/2014	3	252	
A4	1	10/10/2014	3	252	
A5	1	10/10/2014	3	252	
B3	12	01/11/2008	8	281	
B16	12	25/11/2013	9	253	
B34	12	03/09/2014	11	179	
B37	12	18/01/2012	12	-	
B62	12	03/09/2014	15	-	
B113	12	23/10/2014	18	1328	*mutL* (INDEL)
B114	12	11/04/2012	19	198	
B199	12	18/10/2011	32	1182	
A12	1	14/11/2014	35	179	
C2	4	14/10/2009	36	253	
C3	4	25/02/2010	37	260	
C4	4	25/02/2010	37	-[a]	
C5	4	25/02/2010	37	260[a]	
C6	4	03/03/2010	38	244	
C7	4	23/03/2010	39	244	
C10	4	16/04/2010	40	244	
C8	4	16/04/2010	40	244	

C9	4	16/04/2010	40	244	
C11	4	19/08/2010	41	282	
C12	4	20/08/2010	42	282	
C13	4	20/08/2010	42	27 [b]	
C14	4	20/08/2010	42	27 [b]	
C15	4	20/08/2010	42	27 [b]	
C16	4	20/08/2010	42	- [b]	
C17	4	20/08/2010	42	- [b]	
C18	4	15/04/2011	43	-	*mutL* (STOP)
C20	4	01/07/2011	44	878	
C21	15	14/04/2009	45	- [c]	
C22	15	14/04/2009	45	- [c]	
C23	15	14/04/2009	45	- [c]	
C25	15	20/05/2009	46	253	
C29	15	03/06/2009	48	252	
C30	15	04/06/2009	49	252	
C31	15	25/08/2009	50	- [d]	*mutL* (INDEL)
C32	15	25/08/2009	50	- [d]	
C33	15	25/08/2009	50	- [d]	
C36	15	21/05/2010	52	253	
C42	15	12/07/2010	54	309	
C43	15	28/07/2010	55	108	
C44	15	28/07/2010	55	108	
C45	15	28/07/2010	55	108	
C49	15	22/02/2011	58	395	
C51	15	23/02/2011	59	683	
C54	15	15/06/2011	61	1342	
C55	5	26/06/2009	62	- [e]	
C56	5	26/06/2009	62	- [e]	
C57	5	26/06/2009	62	- [e]	
C58	5	26/06/2009	62	- [e]	
C59	5	26/06/2009	62	- [e]	
C60	5	26/06/2009	62	- [e]	
C61	5	17/11/2009	63	620	
C63	3	02/09/2009	64	27	
C64	3	25/11/2009	65	274	
C65	3	25/11/2009	65	274	
C66	3	25/11/2009	65	274	
C67	3	25/11/2009	65	274	
C68	3	25/11/2009	65	274	
C69	3	12/05/2010	66	-	

C71	9	04/09/2009	67	968	
C73	9	05/11/2010	68	17	
C74	9	03/12/2010	69	1202	*mutL* (STOP)
C76	2	12/05/2009	70	253	
C77	2	03/07/2009	71	308	
C78	2	14/07/2009	72	840	*mutL* (INDEL)
C79	2	14/07/2009	72	620 [f]	*mutL* (STOP)
C80	2	14/07/2009	72	620 [f]	
C81	2	14/07/2009	72	620 [f]	
C82	2	14/07/2009	72	620 [f]	
C83	2	14/07/2009	72	- [f]	
C84	2	14/07/2009	72	620 [f]	
C85	2	06/08/2009	73	- [g]	
C86	2	06/08/2009	73	308	
C87	2	06/08/2009	73	179 [g]	
C88	2	15/12/2009	74	1251	
C89	2	25/03/2010	75	1239	
C91	2	13/01/2011	76	253	
C92	2	01/02/2011	77	252	
C94	10	03/07/2009	78	395	
C95	10	29/07/2009	79	253	
C96	10	29/07/2009	79	253	
C97	10	29/07/2009	79	253	
C98	10	29/07/2009	79	253	
C99	10	29/07/2009	79	253	
C100	10	13/10/2009	80	612	
C101	10	21/10/2009	81	- [h]	
C102	10	21/10/2009	81	- [h]	
C103	10	21/10/2009	81	- [h]	
C104	13	16/05/2009	82	179	
C105	13	25/07/2009	83	840	
C106	13	11/08/2009	84	- [i]	
C107	13	11/08/2009	84	253	
C108	13	11/08/2009	84	179 [i]	
C109	13	11/08/2009	85	840	
C110	13	11/08/2009	85	179	
C111	13	11/08/2009	85	179	
C112	13	11/08/2009	85	179	
C114	13	05/12/2009	86	179	
C115	13	05/12/2009	86	179	
C116	13	04/06/2010	87	871	

C117	13	04/06/2010	87	871	
C118	13	04/06/2010	87	871	
C119	7	23/01/2010	88	-	
C120	7	29/01/2010	89	-	
C123	7	02/04/2010	90	27	*mutS* (INDEL)
C124	7	08/04/2010	91	1753	
C125	7	29/04/2010	92	253	
C126	7	29/04/2010	92	253	
C127	7	29/04/2010	92	164	*mutS* (INDEL)
C128	7	29/04/2010	92	164	
C129	7	29/04/2010	92	871	*mutL* (INDEL)
C131	7	08/05/2010	93	253	
C133	7	28/05/2010	94	253	*mutS* (INDEL)
C134	7	19/12/2010	95	253	*mutS* (INDEL)
C135	14	02/11/2009	96	160	
C137	14	16/04/2010	97	260	
C139	14	08/09/2010	98	2102	
C141	14	01/10/2010	99	2102	
C142	14	11/02/2011	100	252	
C143	8	08/07/2010	101	253	
C144	8	08/07/2010	101	253	
C145	8	08/07/2010	101	253	
C146	8	25/06/2010	102	395	
C147	8	25/06/2010	102	395	
C148	8	25/06/2010	102	395	
C149	8	22/03/2011	103	108	
C150	6	27/08/2010	104	253	
C151	6	03/03/2011	105	1244	
C153	6	07/04/2011	106	155	
C155	11	04/12/2010	107	1211	
C156	16	16/11/2010	108	260	
C158	16	03/12/2010	109	155	
C159	16	09/12/2010	110	260	
C160	16	09/12/2010	111	1244	
C161	16	03/03/2011	112	110	
C164	16	08/12/2010	113	-	
C167	16	21/04/2011	114	296	
C168	16	21/12/2010	115	17	
A19	1	16/12/2014	120	-	
A163	1	19/05/2015	137	146	
A36	1	17/02/2015	137	146	

A46	1	07/04/2015	147	17	
A48	1	07/04/2015	147	17	
A52	1	07/04/2015	147	17	
A53	1	07/04/2015	147	17	
A54	1	07/04/2015	147	17	
A55	1	07/04/2015	147	17	
A56	1	07/04/2015	147	17	
A58	1	07/04/2015	147	17	
A60	1	07/04/2015	147	17	
A70	1	07/04/2015	147	17	
A71	1	07/04/2015	147	17	
A72	1	07/04/2015	147	17	
A73	1	07/04/2015	147	17	
A75	1	07/04/2015	147	17	
A76	1	07/04/2015	147	17	
A100	1	07/04/2015	148	17	
A106	1	07/04/2015	148	175	
A107	1	07/04/2015	148	175	
A77	1	07/04/2015	148	175	
A78	1	07/04/2015	148	17	
A80	1	07/04/2015	148	175	
A81	1	07/04/2015	148	17	
A82	1	07/04/2015	148	17	
A85	1	07/04/2015	148	175	
A86	1	07/04/2015	148	175	
A90	1	07/04/2015	148	175	
A91	1	07/04/2015	148	175	
A92	1	07/04/2015	148	175	
A95	1	07/04/2015	148	175	
A97	1	07/04/2015	148	175	
A119	1	15/05/2015	149	667	
A122	1	15/05/2015	149	667	
A123	1	15/05/2015	149	667	
A126	1	15/05/2015	149	667	
A130	1	15/05/2015	149	667	
A134	1	15/05/2015	149	667	
A137	1	15/05/2015	149	667	
A141	1	15/05/2015	149	667	
A144	1	15/05/2015	149	667	
A147	1	15/05/2015	149	667	
A148	1	15/05/2015	149	667	

A151	1	15/05/2015	149	667	
A154	1	15/05/2015	149	667	
A156	1	15/05/2015	149	667	

Supplementary Table S2. MLST profiles of isolates where incomplete profiles were obtained or the MLST profile was novel.

Isolate	ST	Closest ST	*Pseudomonas aeruginosa* MLST loci allele numbers						
			acs	*aro*	*gua*	*mut*	*nuo*	*pps*	*trp*
A19	NF	92 or 261	105	5	30	-	1	4	14
B37	Novel		107	4	3	27	12	7	128
B62	NF	1404	16	-	6	3	4	7	1
C101	NF	303 (4 loci)	16	-	12	18	3	4	9
C102	NF	304 (4 loci)	16	-	12	18	3	4	9
C103	NF	305 (4 loci)	16	-	12	18	3	4	9
C106	NF	156,179,353,1494 (6 loci)	-	27	28	3	4	13	7
C119	Novel		5	1	109	3	1	1	47
C120	Novel		17	5	11	5	4	29	2
C164	NF	1240,1985 (4 loci)	28	5	46	5	1	-	61
C16	NF	27,120,2314 (5 loci)	6	-	6	113	4	6	7
C17	NF	27,120,2314 (5 loci)	6	-	6	113	4	6	7
C18	NF	158,179,180,1496,2063,2109 (6 loci)	36	27	28	-	4	13	7
C21	Novel		22	6	1	3	1	76	1
C22	Novel		22	6	1	3	1	76	1
C23	Novel		22	6	1	3	1	76	1
C31	NF	155,677,1276 (5 loci)	28	5	36	-	3	13	7

C32	NF	155,677,1276 (5 loci)	28	5	36	-	3	13	7
C33	NF	155,677,1276 (5 loci)	28	5	36	-	3	13	7
C4	NF	260 (6 loci)	14	5	-	7	4	13	7
C55	Novel		28	5	11	11	15	75	1
C56	Novel		28	5	11	11	15	75	1
C57	Novel		28	5	11	11	15	75	1
C58	Novel		28	5	11	11	15	75	1
C59	Novel		28	5	11	11	15	75	1
C60	Novel		28	5	11	11	15	75	1
C69	NF	1707, 2055 (6 loci)	16	24	1	149	4	-	19
C83	NF	620 (6 loci)	9	7	63	13	8	-	8
C85	NF	156, 179, 353, 1494 (6 loci)	-	27	28	3	4	13	7

Table S3. Clonal lineages isolated from multiple patients within individual centres. Sets of isolates from different patients attending the same centre are grouped by clonal lineage. For each such group, whole genome pairwise comparisons were carried out to determine the number of variant SNPs and INDELs. *These isolates share a novel MLST profile and they cluster according to SNP phylogeny.

Isolate	Centre	Isolation date	Patient	MLST	Comparison	SNPs	INDELs
C6	4	03/03/2010	38	244	C6-C7	179	8
					C6-C8	3790	79
C7	4	23/03/2010	39	244	C6-C9	3733	74
					C6-C10	3833	62
C8	4	16/04/2010	40	244	C7-C8	3736	82
					C7-C9	3714	75
C9	4	16/04/2010	40	244	C8-C9	281	5
					C7-C10	3929	65
C10	4	16/04/2010	40	244	C8-C10	603	14
					C9-C10	515	11
C11	4	19/08/2010	41	282	C11-C12	340	8
C12	4	20/08/2010	42	282			
C29	15	03/06/2009	48	252	C29-C30	168	3
C30	15	04/06/2009	49	252			
C25	15	20/05/2009	46	253	C25-C36	3428	43
C36	15	21/05/2010	52	253			
C91	2	13/01/2011	76	253	C76-C91	846	5
C76	2	12/05/2009	70	253			
C77	2	03/07/2009	71	308	C77-C86	277	1
C86	2	06/08/2009	73	308			
C105	13	25/07/2009	83	840	C105-C109	131	8
C109	13	11/08/2009	85	840			
C104	13	16/05/2009	82	179	C104-C108	10,551	114
					C104-C110	2863	39
C108	13	11/08/2009	84	179	C104-C111	2765	52
					C104-C112	1913	25
C110	13	11/08/2009	85	179	C104-C114	6795	96
					C104-C115	6972	105
					C108-C110	4780	78
					C108-C111	4852	66
					C108-C112	3780	62

C111	13	11/08/2009	85	179	C108-C114	10,963	136
					C108-C115	10,833	133
					C110-C111	176	8
C112	13	11/08/2009	85	179	C110-C112	198	7
					C110-C114	3115	40
					C110-C115	3107	45
C114	13	05/12/2009	86	179	C111-C112	148	4
					C111-C114	2838	43
					C111-C115	2789	57
C115	13	05/12/2009	86	179	C112-C114	1941	29
					C112-C115	1895	31
					C114-C115	281	4
C125	7	29/04/2010	92	253	C125-C126	736	27
					C125-C131	1159	22
C126	7	29/04/2010	92	253	C125-C133	872	22
					C125-C134	817	21
C131	7	08/05/2010	93	253	C126-C131	9636	99
					C126-C133	5847	71
C133	7	28/05/2010	94	253	C126-C134	5883	70
					C131-C133	3486	40
C134	7	19/12/2010	95	253	C131-C134	3400	38
					C133-C134	330	4
C156	16	16/11/2010	108	260	C156-C159	160	3
C159	16	09/12/2010	110	260			
C139	14	08/09/2010	98	ST2102	C139-C141	177	3
C141	14	01/10/2010	99	ST2102			

Supplementary Table S4 and Figure S1 are provided in additional files:

Table S4. Clone-specific deletions, relative to PAO1

Figure S1. Core genome SNP phylogeny showing the distribution of bronchiectasis isolates. The figure shows analysis of the genomes of all bronchiectasis isolates used in this study (highlighted in blue) alongside 331 genomes from Kos et al. [14] and the genomes of commonly studied strains PAO1 (labelled PAO1107), PA14 (UCBPPPA14109), PA7 and LESB58. Line colours indicate the two major clusters of *P. aeruginosa* (I, green; II, blue) as well as those isolates not clustering in the two main groups (red).

Figure S2. Distribution of loss of function mutations. For the genes listed in Table 2, where the number of independent occurrences of a loss of function mutation was equal to or greater than five, the Figure indicates the number of isolates carrying mutations in 0, 1, 2, 3, 4, 5 or 6 of these genes.

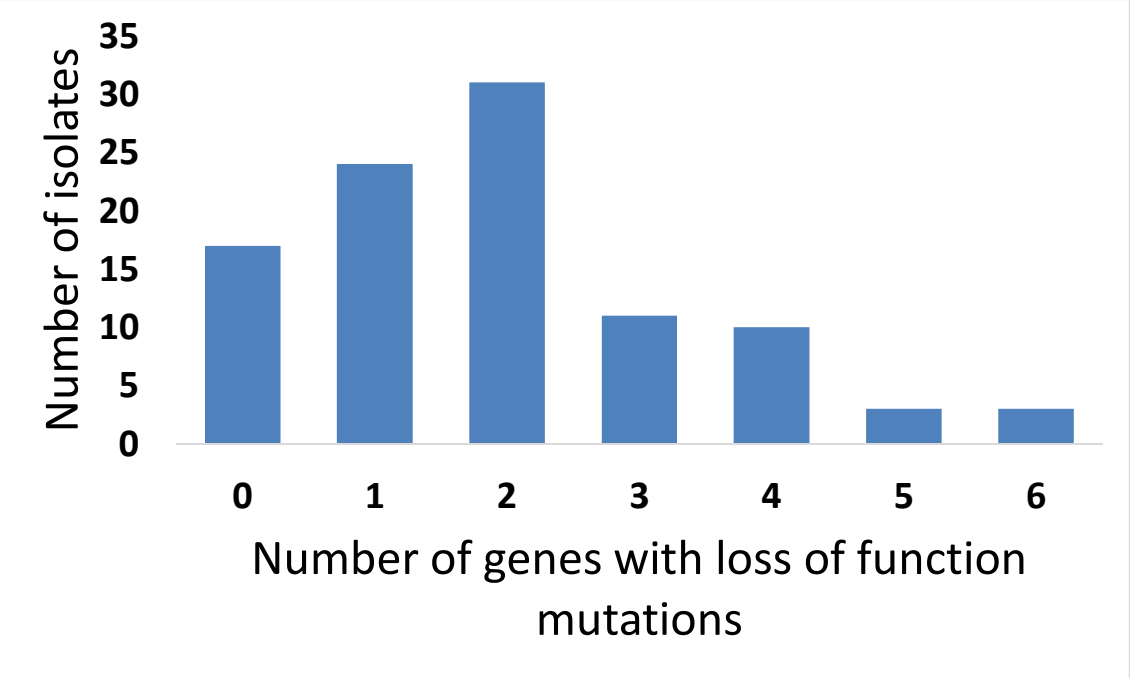


Table S5. Full list of loss of function mutations identified by variant calling. Available via the Figshare link https://figshare.com/s/ff426bae75ee64804aa1

Table S6. Loss of function mutations present in each bronchiectasis isolate genome. For the genes listed in Table 2, where the number of independent occurrences of a loss of function mutation was equal to or greater than five, the genes carrying such mutations are shown for each of the bronchiectasis isolates. 17 of the 99 isolates carried none of these mutations. It is worth noting that isolates found co-infecting individual patients did not share the same mutation profile (C125, C127

and C129 in patient 92; C12 and C13 in patient 42; C78 and C79 in patient 72; C86 and C87 in patient 73; C107 and C108 in patient 84; C109 and C110 in patient 85; A100 and A106 in patient 148).

Isolate	Number of Table2 genes with loss of function mutations	Mutations					
C125	6	PA4469	mexS	ladS	rbdA	betT2	mexA
C74	6	PA4469	mexS	chpA	mutL	pchE	mexB
C87	6	PA4469	bifA	pchE	ladS	chpA	mexA
C123	5	rbdA	PA4469	pchE	mexS	oprM	
C116	5	PA4469	mucA	mexB	oprF	pchE	
C88	5	mucA	pcoA	betT2	oprF	mexA	
C100	4	rbdA	PA4469	mucA	chpA		
C79	4	rbdA	PA4469	mucA	mutL		
C133	4	rbdA	mucA	pchE	oprM		
C95	4	rbdA	mucA	oprM	pcoA		
B16	4	rbdA	mucA	mexB	fimV		
C21	4	PA4469	mucA	mexB	pcoA		
C31	4	PA4469	mucA	mutL	fimV		
C110	4	PA4469	mucA	oprF	mexA		
C108	4	mexB	chpA	mucA	betT2		
C129	4	mucA	mexB	oprF	gmd		
C134	3	rbdA	mucA	oprM			
C143	3	rbdA	mucA	mexB			
C2	3	rbdA	mexB	chpA			
A119	3	PA4469	oprF	pcoA			
A46	3	PA4469	mucA	oprF			
C63	3	betT2	ladS	fimV			
C114	3	mexB	pchE	betT2			
A163	3	mucA	mexB	gmd			
C149	3	mucA	pchE	oprM			
C55	3	mucA	pilJ	oprF			
B113	3	pilJ	mutL	pcoA			
C124	2	PA0054	mucA				
C146	2	PA0054	mexA				
C49	2	PA0054	betT2				
C51	2	PA0054	rbdA				
B114	2	rbdA	mexB				
C25	2	rbdA	mexB				
C36	2	rbdA	mucA				
C86	2	rbdA	mexA				

C20	2	rbdA	pilJ
C137	2	rbdA	PA4469
C6	2	PA4469	mexS
C7	2	PA4469	mexS
C135	2	mucA	mexS
C61	2	PA4469	mexA
C42	2	PA4469	oprF
A100	2	PA4469	fimV
B62	2	PA4469	pilJ
C131	2	betT2	ladS
B34	2	bifA	betT2
C12	2	bifA	betT2
C11	2	bifA	betT2
A1	2	bifA	ladS
B3	2	fimV	pcoA
C167	2	mexA	gmd
C3	2	pilJ	mexB
A2	2	mucA	mexB
C153	2	mexB	oprF
C158	2	mexB	oprF
C161	2	mucA	pcoA
C18	2	oprM	mutL
C155	2	pilJ	oprM
C94	1	PA0054	
C120	1	rbdA	
A3	1	rbdA	
C77	1	rbdA	
C91	1	rbdA	
C69	1	rbdA	
C156	1	rbdA	
C159	1	rbdA	
C107	1	betT2	
C119	1	bifA	
A106	1	bifA	
C43	1	bifA	
B199	1	gmd	
C89	1	gmd	
C168	1	gmd	
A12	1	mexB	
C54	1	mexB	
C92	1	mucA	
C10	1	mucA	
C73	1	mucA	
C142	1	mucA	
C76	1	mucA	
B37	1	mucA	

C101	1	oprM

Table S7. The presence or absence of virulence genes amongst the non-CF BE isolate genomes. Available via the Figshare link https://figshare.com/s/626a59cfd94b13e5cf71

References

1. Martin M. Cutadapt removes adapter sequences from high-throughout sequencin reads. *EMBnet J* 2011: 17(1): 10-12.
2. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 2015: 31(4): 587-589.
3. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013: 29(8): 1072-1075.
4. Laing C, Buchanan C, Taboada EN, Zhang Y, Kropinski A, Villegas A, Thomas JE, Gannon VP. Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. *BMC bioinformatics* 2010: 11: 461.
5. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 2013: 30(12): 2725-2729.
6. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 2007: 23(1): 127-128.
7. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. *Genome biology* 2004: 5(2): R12.
8. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV. Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen. *Nature* 2000: 406(6799): 959-964.
9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009: 25(14): 1754-1760.
10. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009: 25(16): 2078-2079.
11. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 2010: 20(9): 1297-1303.
12. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 2012: 6(2): 80-92.
13. Mowat E, Paterson S, Fothergill JL, Wright EA, Ledson MJ, Walshaw MJ, Brockhurst MA, Winstanley C. Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections. *AmJRespirCrit Care Med* 2011: 183(12): 1674-1679.
14. Kos VN, Deraspe M, McLaughlin RE, Whiteaker JD, Roy PH, Alm RA, Corbeil J, Gardner H. The resistome of Pseudomonas aeruginosa in relationship to phenotypic susceptibility. *Antimicrobial agents and chemotherapy* 2015: 59(1): 427-436.

Table S4. Clone-specific deletions relative to PAO1.

Isolate*	Deleted DNA†		Size of	Deleted Genes‡	
	From	To	deletion (b	From	To
A12	2448472	2723305	274833	PA2226	PA2427
A119	2443094	2624360	181189	PA2221	PA2373
A119	5455322	5497400	42078	PA4857	PA4900
A163	2543016	2705463	162447	PA2305	PA2421
A19	2420900	2721500	300600	PA2201	PA2425
A78	2439040	2574546	135506	PA2218	PA2333
B34	1979190	2015165	35975	PA1820	PA1856
B62	2461847	2686720	224873	PA2237	PA2402
C100	2443095	2724005	280910	PA2221	PA2428
C101	2452047	2578374	126327	PA2229	PA2335
C119	2184340	2247131	62791	PA1997	PA2053
C119	2487183	2689945	202762	PA2258	PA2406
C125	2439064	2729482	290418	PA2218	PA2432
C135	2443095	2588618	145523	PA2221	PA2343
C137	2453340	2560360	107020	PA2231	PA2321
C155	2440000	2585490	145490	PA2218	PA2431
C156	2439060	2475845	36785	PA2218	PA2249
C164	2248926	2451500	202574	PA2055	PA2228
C164	2793790	2810695	16905	PA2475	PA2494
C164	3914150	3930740	16590	PA3497	PA3514
C21	2439057	2713422	274365	PA2218	PA2424
C22	2209250	2442066	232816	PA2018	PA2220
C4	111095	211425	100330	PA0091	PA0781
C4	2439055	2666000	226945	PA2218	PA2400
C44	2443095	2615075	171980	PA2221	PA2365
C5	3841807	3873962	32155	PA3434	PA3463
C51	2071372	2199654	128282	PA1900	PA2010
C54	762796	789380	26584	PA0691	PA0717
C54	2410960	2478550	67590	PA2191	PA2251
C54	2628400	2817536	189136	PA2377	PA2500
C61	4005650	4016519	10869	PA3573	PA3584
C6	2448920	2689471	240551	PA2227	PA2405
C73	2460955	2584136	123181	PA2236	PA2399
C87	2461666	2606428	144762	PA2237	PA2359
C87	2787221	2828217	40996	PA2469	PA2511
C96	2705835	2863924	158089	PA2422	PA2535

*For identical strains with the same deletion, only one strain is listed

†Coordinates of P. aeruginosa PAO1

‡P. aeruginosa strain PAO1 locus ID

REVISED MANUSCRIPT

Title: *Pseudomonas aeruginosa* adaptation and diversification in the non-Cystic Fibrosis bronchiectasis lung

Authors:

Yasmin Hilliam[1$], Matthew P. Moore[1$], Iain L. Lamont[2], Diana Bilton[3], Charles S. Haworth[4], Juliet Foweraker[4], Martin J. Walshaw[5], David Williams[1,6], Joanne L. Fothergill[1], Anthony De Soyza[7]*, Craig Winstanley[1]*

Affliliations:

1. Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

2. Department of Biochemistry, University of Otago, New Zealand

3. Royal Brompton Hospital, London, UK

4. Cambridge Centre for Lung Infection, Papworth Hospital, Cambridge, UK

5. Liverpool Heart and Chest Hospital, Liverpool, UK

6. Institute of Integrative Biology, University of Liverpool, Liverpool, UK

7. Institute for Cellular Medicine, Newcastle University, Newcastle-upon-Tyne, UK

8. Adult Bronchiectasis Service, Freeman Hospital, Newcastle-upon-Tyne, UK

*$ contributed equally

Corresponding author:

Professor Craig Winstanley

Institute of Infection and Global Health,

University of Liverpool,

Ronald Ross Building,

8 West Derby Street,

Liverpool L69 7BE, UK

Email: C.Winstanley@liv.ac.uk

Summary: In bronchiectasis, *P. aeruginosa* co-infections occur; bacterial populations both adapt and diversify by mutation

Abstract

To characterise *Pseudomonas aeruginosa* populations during chronic lung infections of non-Cystic Fibrosis bronchiectasis patients, we used whole genome sequencing to (i) assess the diversity of *P. aeruginosa* and the prevalence of multi-lineage infections, (ii) seek evidence for cross-infection or common source acquisition and (iii) characterize *P. aeruginosa* adaptations.

189 isolates, obtained from the sputa of 91 patients attending 16 adult UK bronchiectasis centres, were whole genome sequenced.

Bronchiectasis isolates were representative of the wider *P. aeruginosa* population. Of 24 patients where multiple isolates were examined, there were seven examples of multi-lineage infections, most likely arising from multiple infection events. The number of nucleotide variants between genomes of isolates from different patients was in some cases similar to the variations observed between isolates from individual patients implying the possible occurrence of cross infection or common source acquisition.

Our data indicate that during infections of bronchiectasis patients, *P. aeruginosa* populations adapt by accumulating loss of function mutations, leading to changes in phenotypes including different modes of iron acquisition and variations in biofilm-associated polysaccharides. The within-population diversification suggests that larger scale longitudinal surveillance studies will be required to capture cross infection or common source acquisition events at an early stage.

Abstract word count: 196

Introduction

Bronchiectasis is a chronic, progressive respiratory disease associated with irreversible widening of the bronchi [1]. Recent data suggest that in the UK incidence rates in women and men have risen to 35.2 and 26.9 respectively per 100,000 person-years [2]. In the USA the prevalence of adult bronchiectasis has been estimated at 52 in 100 000 people, with higher prevalence among women and older individuals [3]. Persistent *Pseudomonas aeruginosa* lung infections of bronchiectasis patients, occuring in approximately 30% of cases, are associated with poorer outcomes and premature mortality [4, 5].

The study of chronic *P. aeruginosa* lung infections has focused on cystic fibrosis (CF)-associated bronchiectasis, where patients are diagnosed, monitored and subjected to antibiotic therapy from a very early age. This contrasts with non-CF bronchiectasis patients, who present at a much older age and often have a shorter history of therapeutic interventions. Hence, bacterial isolates from non-CF bronchiectasis patients exhibit less resistance to antibiotics compared to isolates from adult CF patients [6]. Previous studies have characterized the evolution of *P. aeruginosa* during chronic lung infections in CF patients [7, 8]. More recently, high-resolution analyses have revealed extensive heterogeneity within *P. aeruginosa* populations in the CF lung [9-12], including the co-existence of multiple divergent lineages [13].

In CF, a number of transmissible strains of *P. aeruginosa* have been identified, leading to the introduction of measures to control cross infection [14]. The study of *P. aeruginosa* in relation to non-CF bronchiectasis is less advanced. In our single centre study of 50 *P. aeruginosa* isolates from 40 bronchiectasis patients using molecular typing, there was no compelling evidence for cross infection or a dominant clone [15]. However, whole genome sequence analysis of multiple bronchiectasis isolates has not been carried out. Here, we report the use of genomics to assess the diversity of *P. aeruginosa* strains causing

infections in non-CF bronchiectasis across multiple UK centres, to identify multi-strain infections, and to look for evidence for cross-infection or common source acquisition. We also characterise adaptive mutations and present evidence for within-population divergence during *P. aeruginosa* chronic lung infections of bronchiectasis patients.

Methods

Patients and bacterial isolates

The 189 *P. aeruginosa* isolates used in this study (see Table S1 in the online supplementary material) were isolated from sputum samples obtained from 93 patients with bronchiectasis and chronic *P. aeruginosa* infection (defined as two or more positive respiratory tract cultures in the preceding 12 months) attending 16 adult bronchiectasis centres throughout England and Wales. These included isolates collected as part of a multi-centre nebulized antibiotic trial [16], where patients were enrolled within 21 days of completing a course of antipseudomonal antibiotics for an exacerbation. Additional isolates from Newcastle (n = 8) and Liverpool (n = 53) were collected during observational studies. The methodology used for isolating *P. aeruginosa* from patient sputum samples is described in the online supplementary material.

For 24 patients, sets of isolates (two or more) from the same sample were analysed to look for evidence of multi-lineage infections. For three of these patients (patients 147 – 149), sets of 14 or 15 isolates from a single sample were sequenced for higher resolution analysis of within-population heterogeneity. For some analyses, to avoid biases arising from inclusion of multiple clonal genomes from the same patient, a subset of 99 genomes from 91 patients was used. This subset consisted of one randomly selected genome per clonal lineage per patient (see Table S1 in the online supplementary material). We use the term "clonal lineage" to describe isolates with shared multilocus sequence type (MLST) profile and

clustering according to core genome single nucleotide polymorphism (SNP)-based phylogeny.

DNA preparation and whole genome sequencing

Details of the extraction of genomic DNA from *P. aeruginosa* isolates, library preparation and whole-genome shotgun sequencing using Illumina short read sequencing technology are given in the online supplementary material. The European Nucleotide Archive accession number for the study is PRJEB14952.

Methods used for genome sequence assembly, extraction of MLST data, phylogenetic reconstruction using the core genome, and variant calling by mapping to the genome of PAO1 [17] to identify single nucleotide polymorphism (SNP) or small insertions or deletions (INDELs) are described in online supplementary material.

Identification of large deletions and virulence factor genes

Genome sequences were aligned to the reference genomes *P. aeruginosa* PAO1 (NC_002516, [17]) and *P. aeruginosa* LESB58 (FM209186; [18]) and large clone-specific deletions (10 kb and above) were identified using BRIG [19]. The boundaries of deletions were determined by aligning the genome sequences with the *P. aeruginosa* PAO1 genome using Mauve [20], implemented as part of the Geneious package (www.geneious.com). The presence and absence of virulence factor genes in genome assemblies was determined using Blastable (https://www.github.com/bawee/blastable). The *Pseudomonas* genome database (beta.pseudomonas.com) [21] was used to facilitate analysis of gene function.

Results

Diversity of *P. aeruginosa* Non-CF BE isolates and evidence for *P. aeruginosa* multi-lineage co-infections

Core genome SNP phylogenetic analysis alongside a collection of 331 *P. aeruginosa* isolate genomes from diverse clinical sources [22], indicated that the bronchiectasis isolates were widely distributed (see Figure S1 in the online supplementary material). From the 189 isolates, it was possible to extract complete MLST profiles for 160 (see Tables S1 and S2 in the online supplementary material), with the most widespread sequence types (STs) being ST-253 (PA14-like [23], 14 patients, 8 centres), ST-179 (7 patients, 4 centres), ST-17 (Clone C [23], 5 patients, 3 centres), ST-252 (4 patients, 4 centres) and ST-260 (4 patients, 3 centres). Using core genome SNP phylogeny, previous studies have sub-divided the wider *P. aeruginosa* population into two major groups (group I, which includes strain PAO1, and group II, which includes strain PA14) and one minor group of mostly unrelated clonal lineages [24, 25]. Of a subset of 99 genomes consisting of one randomly selected genome per clonal lineage per patient, 71 were located in group I and 27 in group II (Figure 1). Based on a combination of MLST genotype and core genome SNP phylogeny, of the 24 patients from whose samples multiple isolates were examined, there were seven examples of multi-lineage infections. In one patient (patient 92), three distinct clonal lineages of *P. aeruginosa* were identified. In patients 42, 72, 73, 84, 85 and 148 there were two co-existing lineages (Figure 1).

Evidence for shared lineages causing infections in different patients attending the same centre

The core genome SNP phylogeny identified a number of examples where closely-related clonal lineages were isolated from more than one patient attending the same centre (see Table S3 in the online supplementary material). In order to obtain a higher resolution comparison,

these isolates were analysed using pairwise comparisons across their entire genomes (Table S3), identifying five instances where the genomes of isolates from different patients attending the same centre varied at fewer than 200 sites (C6/C7, C29/C30, C105/109, C139/C141, C156/C159; Figure 2). This level of genome similarity is greater than in some pairwise comparisons of contemporary isolates of the same lineage from the same sputum sample (see Table S3 in the online supplementary material; from 184 variant sites [C110/C111] to >750 variant sites [C125/C126]).

The draft genome sequences of the sub-set of 99 bronchiectasis isolates were examined for the presence of large (> 10kb) deletions. A total of 36 different deletions (25 over 100 kb), ranging in size from 11 to 300 kb and representing independent genetic events, were identified (see Table S4 in the online supplementary material). These were distributed across 28 genomes in the 99-member genome subset. Most genomes had only one deletion, although two (C54 and C164) had three deletions and four (A119, C4, C85 and C119) had two. In most cases, isolates of the same clonal lineages from the same patient shared the same deletions. However, in patients 45, 55, 79 and 92 not all isolates of the same lineage had the same deletion. The genomes of isolate pairs C6/C7, C29/C30, C105/109, C139/C141 and C156/C159, which are from different patients but vary at fewer than 200 sites (Table 1), were indistinguishable by BRIG analysis (example shown in Fig. 3B).

Genomic diversity of isolates within patients can be similar to diversity between patients

In order to further assess the within-patient diversification exhibited by *P. aeruginosa* populations, larger sets of isolates from single sputum samples were analysed for three patients : 147 (15 isolates), 148 (15 isolates) and 149 (14 isolates) (Table 1). For two of these patients, the *P. aeruginosa* population was comprised of a single clonal lineage. For patient 148, two distinct clonal lineages were identified and these two sets of isolates were analysed

separately. In all four isolate sets analysed, the maximum pairwise SNP variations between two isolates of the same lineage exceeded 300, with a median of 179 or greater (Table 1), indicating the occurrence of within-patient diversification.

Loss of function mutations and deletions identified in multiple isolates

We used variant calling approaches to identify independent occurrences of loss of function mutations amongst the sub-set of 99 bronchiectasis isolate genomes. This yielded a number of examples of genes with known functions carrying independent loss of function mutations in multiple isolates (Table 2; see Table S5 in the online supplementary material). These include genes linked to mucoidy, virulence, osmoprotection, biofilm formation, motility, DNA repair and antimicrobial resistance (Table 2). The genes encoding all three components of the MexAB-OprM efflux pump appear amongst the most common loss of function mutations. Multiple isolates also carried loss of function mutations in genes encoding regulators (including *lasR*, *algU*, *fleR*, *vfr*). Among the 99 bronchiectasis isolates, the number of genes with loss of function mutations as listed in Table 2 ranged from 0 to 6 (see Figure S2 and Table S6 in the online supplementary material).

Hypermutability is a common trait amongst CF isolates of *P. aeruginosa*. Of the 99 panel isolates, 11 carried loss of function mutations in the DNA mismatch repair genes, *mutS* or *mutL* (see Table S1 in the online supplementary material). All but two of these were confirmed as having the hypermutable phenotype.

An alignment of all of the genomes containing deletions >10 kb relative to the genome of strain PAO1 revealed a strikingly non-random distribution, with 30 of the 36 deletions lying within the 1.9 to 2.8 Mb portion of the strain PAO1 genome. Genes within

this region include the *psl* genes, encoding an extracellular polysaccharide [26], genes encoding the siderophore pyoverdine, and genes encoding a type VI secretion apparatus [27].

We next specifically examined one representative of each of the 99 clonal lineages for the presence or absence of genes associated with pathogenicity (see Table S6 in the online supplementary material). 23 of these genomes lacked one or more of the *psl* genes. In contrast, all of the genomes contained all of the *alg* genes required for making alginate and the *pel* genes required for making Pel exopolysaccharide. Eleven of the genomes lacked genes required for synthesis of pyoverdine, with nine of these also lacking an *fpvA* receptor gene for uptake of ferripyoverdine, although the genes required for synthesis of an alternative siderophore, pyochelin, were present in all cases. Eleven of the genomes also lacked two or more genes of the Type VI secretion system [PA2360 (*hsiA3*) – PA2373 (*vgrG3*)] (see Table S6 in the online supplementary material). These findings are consistent with the occurrence of deletions of the region of the genome containing Psl, pyoverdine and type VI secretion genes in multiple isolates, although in some isolates smaller deletions (< 10kb) were detected.

Discussion

We used whole genome sequencing to obtain a cross section of the diversity of *P. aeruginosa* strains causing infections in bronchiectasis in the UK. Our data suggest that the distribution of *P. aeruginosa* lineages found amongst the bronchiectasis isolate collection broadly represents what is present in the global *P. aeruginosa* population. In contrast to CF [14], we found no data to suggest that there is a widespread transmissible strain amongst the UK non-CF bronchiectasis community. However, our study did not include large numbers of patients from individual centres. Lineages such as PA14-like and Clone C, that are naturally more abundant in nature [23], were also amongst the most abundant in the bronchiectasis

collection. Because some lineages are naturally more abundant, their occurrence (based on MLST) in multiple patients is not necessarily indicative of common source or cross infection. Whole genome sequencing offers higher resolution than methods such as MLST, allowing us to address this issue.

In a previous comparison of paired isolates from patients within the same bronchiectasis centre, in most patients (9 of 10) the two isolates shared a common genotype, with one patient found to be infected with two strains simultaneously [15]. In this study, of 24 patients from whose samples multiple isolates were examined, seven had multi-lineage infections. Similar multi-lineage infections have also been reported in CF, generally associated with children [28]. A number of studies in CF have also demonstrated the phenotypic [9, 11, 12] and genomic [10, 13, 29] diversification of single lineage *P. aeruginosa* populations in the CF lung. Here, we show for the first time that similar diversification occurs during infections of non-CF bronchiectasis patients. Both the prevalence of multi-lineage infections, and the diversification that occurs during the infection process emphasise the need to be cautious when interpreting the analysis of sputum samples based on single isolates of *P. aeruginosa*.

We found several examples of isolates from patients attending the same centre, that not only shared the same clonal lineage, but also differed genomically by <200 sites. Genomic variations between isolates from the same patient sample revealed similar, and in some cases higher, levels of variation. The occurrence of isolates with very high genetic relatedness in different patients strongly implies that there has been common source acquisition or cross infection. The extent of the nucleotide variations differentiating two isolates will be dependent upon (i) the length of time since the transmission event and (ii) the rate of mutation of the *P. aeruginosa* population during the infection. Further studies will be

needed to better define the role of cross infection or common source acquisitions in this patient group.

There was clear evidence for bacterial adaptation to the lung environment by the accumulation of mutations and deletions, including loss of function mutations in genes identified previously as being commonly mutated in CF, such as *mucA* (mucoidy) and *lasR* (quorum sensing). It is worth noting, however, that mutations in genes encoding some of the regulators highlighted in previous CF studies (*mexT*, *retS*, *exsD*, *ampR*) were observed either infrequently (two *mexT* and two *ampR* mutants) or not at all (see Table S5 in the online supplementary material). Mutations in global regulators potentially affect numerous processes. In CF, the pathoadaptive genes identified in different studies have varied, suggesting that there are multiple routes to adaptation to the CF lung [7, 8], a scenario which is likely to apply also to non-CF bronchiectasis.

Loss of function mutations in genes encoding the MexAB-OprM efflux pump were common amongst the bronchiectasis isolates. Although generally thought of as a multidrug efflux system important for antibiotic resistance, this system has also been implicated in virulence [30]. Hence, although it may seem counterintuitive that *P. aeruginosa* should adapt by losing an antibiotic resistance-related efflux pump, it may be that the driver for selection is related to a function other than antibiotic efflux. In contrast, the loss of function mutations in *mexS* can be linked directly to antibiotic resistance, since mutations in *mexS* promote upregulation of the MexEF-OprN MDR efflux pump [31].

The prevalence amongst non-CF BE isolates of deletions in a specific genomic region encoding pyoverdine and Psl polysaccharide was higher than in a dataset of 331 *P. aeruginosa* clinical isolate genomes [22], where 22 genomes lacked one or more *psl* genes, only three lacked one or more of the pyoverdine synthesis genes, and only one did not have an *fpvA* receptor gene. *P. aeruginosa* can utilise multiple pathways for iron acquisition [32].

During chronic lung infections in CF, *P. aeruginosa* adapts by favouring the heme utilisation route for iron acquisition rather than the pyoverdine siderophore system [33]. Our observations suggest a similar adaptation in non-CF bronchiectasis.

In order to protect itself from hostile environmental conditions or host defences *P. aeruginosa* can produce three exopolysaccharides contributing to biofilm formation: alginate, Psl and Pel [26]. It has been suggested that Psl is a key surface attachment determinant [34], whereas in the CF lung free-floating biofilm structures may be more important [35]. Other mutations favouring the production of Pel rather than Psl include mutations in *bifA* [36], *rbdA* [37], *oprF* [38] and *ladS* [39]. Hence, overall our observations indicate that in non-CF BE chronic lung infections, the Pel and alginate exopolysaccharides are favoured over Psl.

Other common loss of function mutations (in *pilJ*, *chpA* and *fimV*) are implicated in lost or amended twitching motility, an adaptation also seen both in CF [8] and in an artificial sputum biofilm model [40], suggesting that this may be an adaptation related to the viscosity of the sputum environment.

Our study represents the first comparative genomics analysis of multiple *P. aeruginosa* isolates associated with chronic lung infections of non-CF bronchiectasis patients. Although a larger, more targeted study, analysing greater numbers of isolates per sample, would be needed to determine the true prevalence of multi-lineage infections, this observation does suggest that it is common for multiple *P. aeruginosa* lineages to co-exist in bronchiectasis infections. Our study also demonstrates that within-sample diversity can be comparable in scale to the genetic variations that occur between isolates from different patients attending the same centre. These observations suggest that there is an urgent need for more detailed and larger scale longitudinal studies in non-CF patients, and for surveillance that captures the diversity within centres and would identify cross infection or common

source acquisition events earlier, allowing measures to be taken in order to minimise the spread of this important pathogen.

Acknowledgements

This work was supported by the UK Medical Research Council as part of the BRONCH-UK grant (Reference MR/L011263/1) awarded to AdS as Principal Investigator, with CW amongst the Co-Investigators. IL would like to acknowledge funding support from Cystic Fibrosis New Zealand, the New Zealand Lotteries Board (Health) and the NHMRC (Australia). We would like to thank the investigators and patients involved in the PROMIS trial [16]. We also thank Paul Roberts (Royal Liverpool and Broadgreen University Hospitals NHS Trust) for technical assistance.

Figure legends.

Figure 1. Evidence for multi-lineage co-infections in seven patients. The figure shows a core genome SNP phylogeny for the sub-set of 99 isolates, confirming that all but one isolate (B113) clusters into one of two major groups. Each bronchiectasis centre is represented by a different colour. Arrows sharing the same colour indicate isolates that were obtained from the same patient. The three isolates from the same patient 92 sample are numbered 1-3.

Figure 2. Example pairwise comparisons between isolates sharing the same clonal lineage that were isolated from more than one patient attending the same centre. The number of SNP variations are indicated, with the number of IN-DEL variations shown in brackets. Full details are shown in Table S3 in the online supplementary material. The five examples where isolates shared fewer than 200 variant sites are highlighted in green. All

isolates of ST-244 from patients attending Centre 4 were compared, with similarity graded according to variant sites (<200, green; 200 – 3000 orange; >3000 red).

Figure 3. Examples of alignment of genomes of bronchiectasis strains with that of reference strain *P. aeruginosa* PAO1. Sequences identified as present (dark grey) or absent (white) in the genome of PAO1 are indicated. (A) Isolates of the same lineage (ST-253) from the same patient. From innermost to outermost, C95, C97, C98, C99, C96. A deletion present in isolate C96 only is highlighted (see arrow). (B) Pairs of isolates (from innermost to outermost C6 and C7; and C156 and C159) that both share the same clonal lineage but are from different patients attending the same hospital. Isolates C6 and C7 share a large deletion and isolates C156 and C159 share a smaller overlapping deletion (Supplementary Table S4), as indicated (see arrows). (C) Isolates of different lineages from the same patient. From innermost to outermost, A77, A80, A85 (all ST-175); A78, A81, A82 (all ST-17). A large deletion present in the ST-17 isolates is indicated by an arrow. The figures were generated using BRIG.

References

1. De Soyza A, Brown JS, Loebinger MR, Bronchiectasis R, Academic N. Research priorities in bronchiectasis. *Thorax* 2013: 68(7): 695-696.
2. Quint JK, Millett ER, Joshi M, Navaratnam V, Thomas SL, Hurst JR, Smeeth L, Brown JS. Changes in the incidence, prevalence and mortality of bronchiectasis in the UK from 2004 to 2013: a population-based cohort study. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 2016: 47(1): 186-193.
3. Weycker D, Edelsberg J, Oster G, Tino G. Association between risk of acute exacerbations and age in patients with bronchiectasis. *Chest* 2005: 128(4): 153s-153s.
4. Finch S, McDonnell MJ, Abo-Leyah H, Aliberti S, Chalmers JD. A Comprehensive Analysis of the Impact of Pseudomonas aeruginosa Colonization on Prognosis in Adult Bronchiectasis. *Ann Am Thorac Soc* 2015: 12(11): 1602-1611.
5. Chalmers JD, Aliberti S, Blasi F. Management of bronchiectasis in adults. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 2015: 45(5): 1446-1462.
6. McDonnell MJ, Jary HR, Perry A, MacFarlane JG, Hester KL, Small T, Molyneux C, Perry JD, Walton KE, De Soyza A. Non cystic fibrosis bronchiectasis: A longitudinal retrospective observational

cohort study of Pseudomonas persistence and resistance. *Respiratory medicine* 2015: 109(6): 716-726.

7. Marvig RL, Sommer LM, Jelsbak L, Molin S, Johansen HK. Evolutionary insight from whole-genome sequencing of Pseudomonas aeruginosa from cystic fibrosis patients. *Future microbiology* 2015: 10(4): 599-611.

8. Winstanley C, O'Brien S, Brockhurst MA. Pseudomonas aeruginosa Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. *Trends in microbiology* 2016: 24: 327-337.

9. Mowat E, Paterson S, Fothergill JL, Wright EA, Ledson MJ, Walshaw MJ, Brockhurst MA, Winstanley C. Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections. *AmJRespirCrit Care Med* 2011: 183(12): 1674-1679.

10. Jorth P, Staudinger BJ, Wu X, Hisert KB, Hayden H, Garudathri J, Harding CL, Radey MC, Rezayat A, Bautista G, Berrington WR, Goddard AF, Zheng C, Angermeyer A, Brittnacher MJ, Kitzman J, Shendure J, Fligner CL, Mittler J, Aitken ML, Manoil C, Bruce JE, Yahr TL, Singh PK. Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. *Cell Host Microbe* 2015: 18(3): 307-319.

11. Darch SE, McNally A, Harrison F, Corander J, Barr HL, Paszkiewicz K, Holden S, Fogarty A, Crusz SA, Diggle SP. Recombination is a key driver of genomic and phenotypic diversity in a Pseudomonas aeruginosa population during cystic fibrosis infection. *Sci Rep* 2015: 5: 7649.

12. Workentine ML, Sibley CD, Glezerson B, Purighalla S, Norgaard-Gron JC, Parkins MD, Rabin HR, Surette MG. Phenotypic heterogeneity of Pseudomonas aeruginosa populations in a cystic fibrosis patient. *PloS one* 2013: 8(4): e60225.

13. Williams D, Evans B, Haldenby S, Walshaw MJ, Brockhurst MA, Winstanley C, Paterson S. Divergent, coexisting Pseudomonas aeruginosa lineages in chronic cystic fibrosis lung infections. *American journal of respiratory and critical care medicine* 2015: 191(7): 775-785.

14. Fothergill JL, Walshaw MJ, Winstanley C. Transmissible strains of Pseudomonas aeruginosa in Cystic Fibrosis lung infections. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 2012.

15. De Soyza A, Perry A, Hall AJ, Sunny SS, Walton KE, Mustafa N, Turton J, Kenna DT, Winstanley C. Molecular epidemiological analysis suggests cross-infection with Pseudomonas aeruginosa is rare in non-cystic fibrosis bronchiectasis. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 2014: 43(3): 900-903.

16. Haworth CS, Foweraker JE, Wilkinson P, Kenyon RF, Bilton D. Inhaled colistin in patients with bronchiectasis and chronic Pseudomonas aeruginosa infection. *American journal of respiratory and critical care medicine* 2014: 189(8): 975-982.

17. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV. Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen. *Nature* 2000: 406(6799): 959-964.

18. Winstanley C, Langille MG, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, Sanschagrin F, Thomson NR, Winsor GL, Quail MA, Lennard N, Bignell A, Clarke L, Seeger K, Saunders D, Harris D, Parkhill J, Hancock RE, Brinkman FS, Levesque RC. Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of Pseudomonas aeruginosa. *Genome Res* 2009: 19(1): 12-23.

19. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics* 2011: 12: 402.

20. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004: 14(7): 1394-1403.

21. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. *Nucleic acids research* 2016: 44(D1): D646-653.
22. Kos VN, Deraspe M, McLaughlin RE, Whiteaker JD, Roy PH, Alm RA, Corbeil J, Gardner H. The resistome of Pseudomonas aeruginosa in relationship to phenotypic susceptibility. *Antimicrobial agents and chemotherapy* 2015: 59(1): 427-436.
23. Fischer S, Klockgether J, Moran Losada P, Chouvarine P, Cramer N, Davenport CF, Dethlefsen S, Dorda M, Goesmann A, Hilker R, Mielke S, Schonfelder T, Suerbaum S, Turk O, Woltemate S, Wiehlmann L, Tummler B. Intraclonal genome diversity of the major Pseudomonas aeruginosa clones C and PA14. *Environ Microbiol Rep* 2016: 8(2): 227-234.
24. Stewart L, Ford A, Sangal V, Jeukens J, Boyle B, Kukavica-Ibrulj I, Caim S, Crossman L, Hoskisson PA, Levesque R, Tucker NP. Draft genomes of 12 host-adapted and environmental isolates of Pseudomonas aeruginosa and their positions in the core genome phylogeny. *Pathog Dis* 2014: 71(1): 20-25.
25. Freschi L, Jeukens J, Kukavica-Ibrulj I, Boyle B, Dupont MJ, Laroche J, Larose S, Maaroufi H, Fothergill JL, Moore M, Winsor GL, Aaron SD, Barbeau J, Bell SC, Burns JL, Camara M, Cantin A, Charette SJ, Dewar K, Deziel E, Grimwood K, Hancock RE, Harrison JJ, Heeb S, Jelsbak L, Jia B, Kenna DT, Kidd TJ, Klockgether J, Lam JS, Lamont IL, Lewenza S, Loman N, Malouin F, Manos J, McArthur AG, McKeown J, Milot J, Naghra H, Nguyen D, Pereira SK, Perron GG, Pirnay JP, Rainey PB, Rousseau S, Santos PM, Stephenson A, Taylor V, Turton JF, Waglechner N, Williams P, Thrane SW, Wright GD, Brinkman FS, Tucker NP, Tummler B, Winstanley C, Levesque RC. Clinical utilization of genomics data produced by the international Pseudomonas aeruginosa consortium. *Frontiers in microbiology* 2015: 6: 1036.
26. Franklin MJ, Nivens DE, Weadge JT, Howell PL. Biosynthesis of the Pseudomonas aeruginosa Extracellular Polysaccharides, Alginate, Pel, and Psl. *Frontiers in microbiology* 2011: 2: 167.
27. Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, Goodman AL, Joachimiak G, Ordonez CL, Lory S, Walz T, Joachimiak A, Mekalanos JJ. A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus. *Science* 2006: 312(5779): 1526-1530.
28. Hall AJ, Fothergill JL, McNamara PS, Southern KW, Winstanley C. Turnover of strains and intraclonal variation amongst Pseudomonas aeruginosa isolates from paediatric CF patients. *Diagnostic microbiology and infectious disease* 2014: 80(4): 324-326.
29. Chung JC, Becq J, Fraser L, Schulz-Trieglaff O, Bond NJ, Foweraker J, Bruce KD, Smith GP, Welch M. Genomic variation among contemporary Pseudomonas aeruginosa isolates from chronically infected cystic fibrosis patients. *Journal of bacteriology* 2012: 194(18): 4857-4866.
30. Hirakata Y, Srikumar R, Poole K, Gotoh N, Suematsu T, Kohno S, Kamihira S, Hancock RE, Speert DP. Multidrug efflux systems play an important role in the invasiveness of Pseudomonas aeruginosa. *JExpMed* 2002: 196(1): 109-118.
31. Sobel ML, Neshat S, Poole K. Mutations in PA2491 (mexS) promote MexT-dependent mexEF-oprN expression and multidrug resistance in a clinical strain of Pseudomonas aeruginosa. *Journal of bacteriology* 2005: 187(4): 1246-1253.
32. Konings AF, Martin LW, Sharples KJ, Roddam LF, Latham R, Reid DW, Lamont IL. Pseudomonas aeruginosa uses multiple pathways to acquire iron during chronic infection in cystic fibrosis lungs. *Infection and immunity* 2013: 81(8): 2697-2704.
33. Marvig RL, Damkiaer S, Khademi SM, Markussen TM, Molin S, Jelsbak L. Within-host evolution of Pseudomonas aeruginosa reveals adaptation toward iron acquisition from hemoglobin. *mBio* 2014: 5(3): e00966-00914.
34. Colvin KM, Irie Y, Tart CS, Urbano R, Whitney JC, Ryder C, Howell PL, Wozniak DJ, Parsek MR. The Pel and Psl polysaccharides provide Pseudomonas aeruginosa structural redundancy within the biofilm matrix. *Environmental microbiology* 2012: 14(8): 1913-1928.

35. Bjarnsholt T, Jensen PO, Fiandaca MJ, Pedersen J, Hansen CR, Andersen CB, Pressler T, Givskov M, Hoiby N. Pseudomonas aeruginosa biofilms in the respiratory tract of cystic fibrosis patients. *PediatrPulmonol* 2009: 44(6): 547-558.

36. Kuchma SL, Brothers KM, Merritt JH, Liberati NT, Ausubel FM, O'Toole GA. BifA, a cyclic-Di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by Pseudomonas aeruginosa PA14. *Journal of bacteriology* 2007: 189(22): 8165-8178.

37. An S, Wu J, Zhang LH. Modulation of Pseudomonas aeruginosa biofilm dispersal by a cyclic-Di-GMP phosphodiesterase with a putative hypoxia-sensing domain. *Applied and environmental microbiology* 2010: 76(24): 8160-8173.

38. Bouffartigues E, Moscoso JA, Duchesne R, Rosay T, Fito-Boncompte L, Gicquel G, Maillot O, Benard M, Bazire A, Brenner-Weiss G, Lesouhaitier O, Lerouge P, Dufour A, Orange N, Feuilloley MG, Overhage J, Filloux A, Chevalier S. The absence of the Pseudomonas aeruginosa OprF protein leads to increased biofilm formation through variation in c-di-GMP level. *Frontiers in microbiology* 2015: 6: 630.

39. Records AR, Gross DC. Sensor kinases RetS and LadS regulate Pseudomonas syringae type VI secretion and virulence factors. *Journal of bacteriology* 2010: 192(14): 3584-3596.

40. Davies E, James CE, Williams D, O'Brien S, Fothergill JL, Haldenby S, Paterson S, Winstanley C, Brockhurst MA. Temperate phages both mediate and drive adaptative evolution in pathogen biofilms. *Proceedings of the National Academy of Sciences of the United States of America* 2016: in press.

Table 1. Summary of genomic diversity observed within the same clonal lineage of *P. aeruginosa* in individual patients

Patient	Number of isolates	Mean SNPs	Median SNPs	SNP range	Mean Indels	Median Indels	Indel range
Patient 147	15	336.35	261.00	88-640	15.20	14.00	0-35
Patient 148 (ST17)	4	451.50	482.50	159-654	23.83	25.50	6-34
Patient 148 (ST175)	11	195.45	179.00	79-403	9.27	5.00	0-36
Patient 149	14	209.01	206.00	68-327	11.40	10.00	3-28

The table indicates the number of single nucleotide polymorphisms (SNPs) and small insertion and deletion (INDEL) differences between the genomes of contemporary isolates from single sputum samples.

Table 2 Loss of function mutations occurring in multiple isolates. Only mutations predicted to lead to loss of function were included (ie. introduction of a stop codon, or a frame-shift mutation). The number of independent mutations indicates the number of isolates carrying unique mutations in the listed gene. The Table shows those genes where the number of independent occurrences of a mutation was equal to or greater than five.

Gene	PAO1 gene number	Number of independent occurrences of a mutation	Function / Comment
mexB	PA0426	16	Transporter from MexAB-OprM efflux pump, antibiotic resistance, virulence
mucA	PA0763	13	Anti-sigma factor, mutations can lead to mucoidy.
betT2	PA5291	9	Transporter, uptake of small molecules such as choline and glycine betaine, contributing to growth via phosphatidyl choline metabolism and osmoprotection
bifA	PA4367	7	Cyclic-di-GMP phosphodiesterase, inversely regulates biofilm formation
mexA	PA0425	7	Membrane fusion protein from MexAB-OprM efflux pump, antibiotic resistance, virulence
pcoA	PA2065	7	Copper resistance
PA4469	PA4469	7	Hypothetical protein encoded by a gene in same operon as and upstream of *sodM* (superoxide dismutase; response to oxidative stress)
rbdA	PA0861	7	Cyclic-di-GMP phosphodiesterase, modulation of biofilm dispersal, negative regulation of Pel production
pilJ	PA0411	6	Methyl accepting chemotaxis receptor-like protein involved in twitching motility and biofilm formation
oprM	PA0427	6	Outer membrane protein from MexAB-OprM efflux pump, antibiotic resistance, virulence
oprF	PA1777	6	Major porin, biofilm formation
chpA	PA0413	5	Chemotaxis-like chemosensory protein involved in twitching motility
fimV	PA3115	5	Peptidoglycan-binding protein, promotes type IV pilin assembly, twitching motility
ladS	PA3974	5	Sensor kinase, implicated in switch between acute and chronic infection
mutL	PA4946	5	Mismatch repair system, DNA repair, mutation can lead to mutator phenotype
gmd	PA5453	5	GDP-mannose 4,6-dehydratase,

mexS	PA2491	5	Mutations promote MexT-dependent *mexEF-oprN* expression and multidrug resistance
pchE	PA4226	5	Pyochelin synthesis
PA0054	PA0054	5	Hypothetical protein

Online Supplementary Material for:

Pseudomonas aeruginosa adaptation and diversification in the non-Cystic Fibrosis bronchiectasis lung

Yasmin Hilliam, Matthew P. Moore, Iain L. Lamont, Diana Bilton, Charles S. Haworth, Juliet Foweraker, Martin J. Walshaw, David Williams, Joanne L. Fothergill, Anthony De Soyza, Craig Winstanley

Supplementary Methods

Isolation of *P. aeruginosa* from sputum samples

All samples were cultured for routine quantitative microbiology. Sputum was homogenised with equal parts of 0.1% (v/v) dithiothreitol, and the homogenised sample was diluted in sterile distilled water to 1 in 200 and 1 in 10,000. Both dilutions were spread by Whitley Automatic Spiral Plater (WASP) onto agar plates. Cultures on Columbia blood agar, chocolate agar and cysteine lactose electrolyte deficient agar (CLED) were incubated at 37^{o}C in 5% (v/v) CO_2 for up to 48 h. Cultures on Pseudomonas agar plus CFC supplement (PCFC) were incubated at 37^{o}C in air for up to 48 h and then checked for small colony variants (SCV) after incubating on the bench for up to a further 48 h. Colony forming unit (CFU) representatives of up to four colonial morphotypes were sub cultured and stored in 15% (v/v) Glycerol for archiving at -80^{o}C.

DNA preparation and genome sequencing

Genomic DNA was extracted from isolates using a Promega Wizard Genomic DNA Purification Kit, quantified using a Qubit 3.0 fluoromiter (Qubit dsDNA broad range assay kit, Life Technologies) and tested for purity using a NanoDrop 1000 spectrophotometer (Thermo Scientific). Library preparation and whole-genome shotgun sequencing was performed by the Centre for Genomic Research at the University of Liverpool, UK using Illumina short read sequencing technology. Shotgun libraries were prepared from the normalised samples using TruSeq Nano library preparation kit. Following library preparation, paired-end sequencing (2 x 100 bp) was performed by multiplexing into one lane of the Illumina HiSeq platform and sequenced with SBS V4 chemistry.

Following processing, the raw Fastq files were trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2.1 [1]. The option –O 3 was used so that the 3′ end of any reads which match the adapter sequence for 3 bp or more were trimmed. The reads were further trimmed using Sickle (https://github.com/najoshi/sickle) version 1.200 with a minimum window quality score of 20. Reads shorter than 10 bp after trimming were removed. If only one read of a pair passed this filter, it was included in the R0 file, with files R1 and R2 containing corresponding paired-end sequences. Quality filtered and adapter trimmed short reads were *de novo* assembled and scaffolded using the A5 MiSeq assembler [2]. Genome assembly quality metrics such as N50, largest contig, and overall number of contigs were produced using QUAST [3].

Core genome SNP phylogeny

The core genome was extracted using Panseq [4] and was defined as 500 bp fragments of all genomes in this study which matched with at least 85% similarity. A phylogenetic tree was approximated from core genome polymorphic sites, not including gaps or ambiguous bases by maximum likelihood with inner node bootstrap (n = 1000) and 10 discrete gamma

categories. All phylogenetic analyses were performed using MEGA6 [5] and visualised using the iTOL software [6]. Long branches were reduced for clarity.

Multilocus Sequence Typing

MLST profiles were extracted based on the pubMLST *Pseudomonas aeruginosa* scheme (http://pubmlst.org/paeruginosa/) using a specific tool (https://github.com/tseemann/mlst). It was not possible to extract complete MLST profiles from all genomes. In the context of this study, a lineage is defined on the basis of MLST profile and core genome SNP phylogeny.

Whole genome pairwise comparisons

Pairwise comparisons between assembled genomes were performed using MUMmer 3.0 [7]. Any positions in the alignment with ambiguous nucleotides were removed.

Read mapping and variant calling

All genome short read files (*fastq*) were mapped to reference genome PAO1 [8] using bwa-075a (mem) [9] producing a sequence alignment map (*sam*) files which were converted to binary alignment map files (*bam*) using picard tools-1.85 (https://broadinstitute.github.io/picard/). The reference genome was first indexed and sorted using bwa and SAMtools [10] respectively and a sequence dictionary created using picard tools-1.85. Variants were called following the Genome Analysis Toolkit (GATK) best practices workflow, as follows: duplicates were marked, the *bam* file indexed and sorted with picard tools-1.85, realignment targets created, INDELs realigned with GATK-3.3 [11] and variants called with the HaplotypeCaller module. Variants were filtered using vcffilter (https://github.com/vcflib/vcflib) with the standard parameters (DP >9, QUAL >10).

Resulting variant call files (*vcf*) were annotated to predict functional outcomes of variants compared with PAO1 genes using SnpEff [12]. Using SAMtools depth any variants within genes to which short reads had not aligned to 100% of its length were excluded from further analysis, leaving substitutions and short insertions or deletions (INDELs) and eliminating genes from functional variant analysis to which reads did not fully align due to sequencing error (lack of coverage) or genuine large deletions/complete absence.

Loss of function mutations

Predicted loss of function mutations were inferred to have been acquired independently in the population if they differed by position or type between genomes. Where the position and type were the same they were inferred to be shared; in genes where a mutation was shared with other genomes, further loss of function mutations were assumed to have been acquired since the common, ancestral acquisition of the shared mutation.

Supplementary Tables:

Supplementary Table S1. Bacterial isolates used in this study. ST refers to the designated multilocus sequence type. *novel shared MLST and they cluster according to SNP phylogeny. Isolates C21-C23 cluster together according to core genome SNP phylogeny. Isolates included in the subset of 99 genomes are highlighted in red. The Mutator column indicates the presence of mutations associated with hypermutability (INDEL, causing a frameshift or STOP, introduction of a stop codon). Of the 11 isolates carrying such a mutation, all but two (B113 and C78) were confirmed as having a hypermutable phenotype using assays reported previously [13]. Isolates labelled with the same letter from [a] to [i] were considered to be from the same lineage as each other on the basis of incomplete MLST profiles (see Table S2) and clustering by core genome SNP phylogeny (see Figure S1).

Isolate ID	Center	Date	Patient	ST	Mutator
A1	1	09/10/2014	1	17	
A2	1	16/10/2014	2	207	
A3	1	10/10/2014	3	252	
A4	1	10/10/2014	3	252	
A5	1	10/10/2014	3	252	
B3	12	01/11/2008	8	281	
B16	12	25/11/2013	9	253	
B34	12	03/09/2014	11	179	
B37	12	18/01/2012	12	-	
B62	12	03/09/2014	15	-	
B113	12	23/10/2014	18	1328	*mutL* (INDEL)
B114	12	11/04/2012	19	198	
B199	12	18/10/2011	32	1182	
A12	1	14/11/2014	35	179	
C2	4	14/10/2009	36	253	
C3	4	25/02/2010	37	260	
C4	4	25/02/2010	37	-[a]	
C5	4	25/02/2010	37	260[a]	
C6	4	03/03/2010	38	244	
C7	4	23/03/2010	39	244	
C10	4	16/04/2010	40	244	
C8	4	16/04/2010	40	244	

C9	4	16/04/2010	40	244	
C11	4	19/08/2010	41	282	
C12	4	20/08/2010	42	282	
C13	4	20/08/2010	42	27 [b]	
C14	4	20/08/2010	42	27 [b]	
C15	4	20/08/2010	42	27 [b]	
C16	4	20/08/2010	42	- [b]	
C17	4	20/08/2010	42	- [b]	
C18	4	15/04/2011	43	-	*mutL* (STOP)
C20	4	01/07/2011	44	878	
C21	15	14/04/2009	45	- [c]	
C22	15	14/04/2009	45	- [c]	
C23	15	14/04/2009	45	- [c]	
C25	15	20/05/2009	46	253	
C29	15	03/06/2009	48	252	
C30	15	04/06/2009	49	252	
C31	15	25/08/2009	50	- [d]	*mutL* (INDEL)
C32	15	25/08/2009	50	- [d]	
C33	15	25/08/2009	50	- [d]	
C36	15	21/05/2010	52	253	
C42	15	12/07/2010	54	309	
C43	15	28/07/2010	55	108	
C44	15	28/07/2010	55	108	
C45	15	28/07/2010	55	108	
C49	15	22/02/2011	58	395	
C51	15	23/02/2011	59	683	
C54	15	15/06/2011	61	1342	
C55	5	26/06/2009	62	- [e]	
C56	5	26/06/2009	62	- [e]	
C57	5	26/06/2009	62	- [e]	
C58	5	26/06/2009	62	- [e]	
C59	5	26/06/2009	62	- [e]	
C60	5	26/06/2009	62	- [e]	
C61	5	17/11/2009	63	620	
C63	3	02/09/2009	64	27	
C64	3	25/11/2009	65	274	
C65	3	25/11/2009	65	274	
C66	3	25/11/2009	65	274	
C67	3	25/11/2009	65	274	
C68	3	25/11/2009	65	274	
C69	3	12/05/2010	66	-	

C71	9	04/09/2009	67	968	
C73	9	05/11/2010	68	17	
C74	9	03/12/2010	69	1202	*mutL* (STOP)
C76	2	12/05/2009	70	253	
C77	2	03/07/2009	71	308	
C78	2	14/07/2009	72	840	*mutL* (INDEL)
C79	2	14/07/2009	72	620 [f]	*mutL* (STOP)
C80	2	14/07/2009	72	620 [f]	
C81	2	14/07/2009	72	620 [f]	
C82	2	14/07/2009	72	620 [f]	
C83	2	14/07/2009	72	- [f]	
C84	2	14/07/2009	72	620 [f]	
C85	2	06/08/2009	73	- [g]	
C86	2	06/08/2009	73	308	
C87	2	06/08/2009	73	179 [g]	
C88	2	15/12/2009	74	1251	
C89	2	25/03/2010	75	1239	
C91	2	13/01/2011	76	253	
C92	2	01/02/2011	77	252	
C94	10	03/07/2009	78	395	
C95	10	29/07/2009	79	253	
C96	10	29/07/2009	79	253	
C97	10	29/07/2009	79	253	
C98	10	29/07/2009	79	253	
C99	10	29/07/2009	79	253	
C100	10	13/10/2009	80	612	
C101	10	21/10/2009	81	- [h]	
C102	10	21/10/2009	81	- [h]	
C103	10	21/10/2009	81	- [h]	
C104	13	16/05/2009	82	179	
C105	13	25/07/2009	83	840	
C106	13	11/08/2009	84	- [i]	
C107	13	11/08/2009	84	253	
C108	13	11/08/2009	84	179 [i]	
C109	13	11/08/2009	85	840	
C110	13	11/08/2009	85	179	
C111	13	11/08/2009	85	179	
C112	13	11/08/2009	85	179	
C114	13	05/12/2009	86	179	
C115	13	05/12/2009	86	179	
C116	13	04/06/2010	87	871	

C117	13	04/06/2010	87	871	
C118	13	04/06/2010	87	871	
C119	7	23/01/2010	88	-	
C120	7	29/01/2010	89	-	
C123	7	02/04/2010	90	27	*mutS* (INDEL)
C124	7	08/04/2010	91	1753	
C125	7	29/04/2010	92	253	
C126	7	29/04/2010	92	253	
C127	7	29/04/2010	92	164	*mutS* (INDEL)
C128	7	29/04/2010	92	164	
C129	7	29/04/2010	92	871	*mutL* (INDEL)
C131	7	08/05/2010	93	253	
C133	7	28/05/2010	94	253	*mutS* (INDEL)
C134	7	19/12/2010	95	253	*mutS* (INDEL)
C135	14	02/11/2009	96	160	
C137	14	16/04/2010	97	260	
C139	14	08/09/2010	98	2102	
C141	14	01/10/2010	99	2102	
C142	14	11/02/2011	100	252	
C143	8	08/07/2010	101	253	
C144	8	08/07/2010	101	253	
C145	8	08/07/2010	101	253	
C146	8	25/06/2010	102	395	
C147	8	25/06/2010	102	395	
C148	8	25/06/2010	102	395	
C149	8	22/03/2011	103	108	
C150	6	27/08/2010	104	253	
C151	6	03/03/2011	105	1244	
C153	6	07/04/2011	106	155	
C155	11	04/12/2010	107	1211	
C156	16	16/11/2010	108	260	
C158	16	03/12/2010	109	155	
C159	16	09/12/2010	110	260	
C160	16	09/12/2010	111	1244	
C161	16	03/03/2011	112	110	
C164	16	08/12/2010	113	-	
C167	16	21/04/2011	114	296	
C168	16	21/12/2010	115	17	
A19	1	16/12/2014	120	-	
A163	1	19/05/2015	137	146	
A36	1	17/02/2015	137	146	

A46	1	07/04/2015	147	17	
A48	1	07/04/2015	147	17	
A52	1	07/04/2015	147	17	
A53	1	07/04/2015	147	17	
A54	1	07/04/2015	147	17	
A55	1	07/04/2015	147	17	
A56	1	07/04/2015	147	17	
A58	1	07/04/2015	147	17	
A60	1	07/04/2015	147	17	
A70	1	07/04/2015	147	17	
A71	1	07/04/2015	147	17	
A72	1	07/04/2015	147	17	
A73	1	07/04/2015	147	17	
A75	1	07/04/2015	147	17	
A76	1	07/04/2015	147	17	
A100	1	07/04/2015	148	17	
A106	1	07/04/2015	148	175	
A107	1	07/04/2015	148	175	
A77	1	07/04/2015	148	175	
A78	1	07/04/2015	148	17	
A80	1	07/04/2015	148	175	
A81	1	07/04/2015	148	17	
A82	1	07/04/2015	148	17	
A85	1	07/04/2015	148	175	
A86	1	07/04/2015	148	175	
A90	1	07/04/2015	148	175	
A91	1	07/04/2015	148	175	
A92	1	07/04/2015	148	175	
A95	1	07/04/2015	148	175	
A97	1	07/04/2015	148	175	
A119	1	15/05/2015	149	667	
A122	1	15/05/2015	149	667	
A123	1	15/05/2015	149	667	
A126	1	15/05/2015	149	667	
A130	1	15/05/2015	149	667	
A134	1	15/05/2015	149	667	
A137	1	15/05/2015	149	667	
A141	1	15/05/2015	149	667	
A144	1	15/05/2015	149	667	
A147	1	15/05/2015	149	667	
A148	1	15/05/2015	149	667	

A151	1	15/05/2015	149	667	
A154	1	15/05/2015	149	667	
A156	1	15/05/2015	149	667	

Supplementary Table S2. MLST profiles of isolates where incomplete profiles were obtained or the MLST profile was novel.

Isolate	ST	Closest ST	*Pseudomonas aeruginosa* MLST loci allele numbers						
			acs	*aro*	*gua*	*mut*	*nuo*	*pps*	*trp*
A19	NF	92 or 261	105	5	30	-	1	4	14
B37	Novel		107	4	3	27	12	7	128
B62	NF	1404	16	-	6	3	4	7	1
C101	NF	303 (4 loci)	16	-	12	18	3	4	9
C102	NF	304 (4 loci)	16	-	12	18	3	4	9
C103	NF	305 (4 loci)	16	-	12	18	3	4	9
C106	NF	156,179,353,1494 (6 loci)	-	27	28	3	4	13	7
C119	Novel		5	1	109	3	1	1	47
C120	Novel		17	5	11	5	4	29	2
C164	NF	1240,1985 (4 loci)	28	5	46	5	1	-	61
C16	NF	27,120,2314 (5 loci)	6	-	6	113	4	6	7
C17	NF	27,120,2314 (5 loci)	6	-	6	113	4	6	7
C18	NF	158,179,180,1496,2063,2109 (6 loci)	36	27	28	-	4	13	7
C21	Novel		22	6	1	3	1	76	1
C22	Novel		22	6	1	3	1	76	1
C23	Novel		22	6	1	3	1	76	1
C31	NF	155,677,1276 (5 loci)	28	5	36	-	3	13	7

C32	NF	155,677,1276 (5 loci)	28	5	36	-	3	13	7
C33	NF	155,677,1276 (5 loci)	28	5	36	-	3	13	7
C4	NF	260 (6 loci)	14	5	-	7	4	13	7
C55	Novel		28	5	11	11	15	75	1
C56	Novel		28	5	11	11	15	75	1
C57	Novel		28	5	11	11	15	75	1
C58	Novel		28	5	11	11	15	75	1
C59	Novel		28	5	11	11	15	75	1
C60	Novel		28	5	11	11	15	75	1
C69	NF	1707, 2055 (6 loci)	16	24	1	149	4	-	19
C83	NF	620 (6 loci)	9	7	63	13	8	-	8
C85	NF	156, 179, 353, 1494 (6 loci)	-	27	28	3	4	13	7

Table S3. Clonal lineages isolated from multiple patients within individual centres. Sets of isolates from different patients attending the same centre are grouped by clonal lineage. For each such group, whole genome pairwise comparisons were carried out to determine the number of variant SNPs and INDELs. *These isolates share a novel MLST profile and they cluster according to SNP phylogeny.

Isolate	Centre	Isolation date	Patient	MLST	Comparison	SNPs	INDELs
C6	4	03/03/2010	38	244	C6-C7	179	8
					C6-C8	3790	79
C7	4	23/03/2010	39	244	C6-C9	3733	74
					C6-C10	3833	62
C8	4	16/04/2010	40	244	C7-C8	3736	82
					C7-C9	3714	75
C9	4	16/04/2010	40	244	C8-C9	281	5
					C7-C10	3929	65
C10	4	16/04/2010	40	244	C8-C10	603	14
					C9-C10	515	11
C11	4	19/08/2010	41	282	C11-C12	340	8
C12	4	20/08/2010	42	282			
C29	15	03/06/2009	48	252	C29-C30	168	3
C30	15	04/06/2009	49	252			
C25	15	20/05/2009	46	253	C25-C36	3428	43
C36	15	21/05/2010	52	253			
C91	2	13/01/2011	76	253	C76-C91	846	5
C76	2	12/05/2009	70	253			
C77	2	03/07/2009	71	308	C77-C86	277	1
C86	2	06/08/2009	73	308			
C105	13	25/07/2009	83	840	C105-C109	131	8
C109	13	11/08/2009	85	840			
C104	13	16/05/2009	82	179	C104-C108	10,551	114
					C104-C110	2863	39
C108	13	11/08/2009	84	179	C104-C111	2765	52
					C104-C112	1913	25
C110	13	11/08/2009	85	179	C104-C114	6795	96
					C104-C115	6972	105
					C108-C110	4780	78
					C108-C111	4852	66
					C108-C112	3780	62

C111	13	11/08/2009	85	179	C108-C114	10,963	136
					C108-C115	10,833	133
					C110-C111	176	8
C112	13	11/08/2009	85	179	C110-C112	198	7
					C110-C114	3115	40
					C110-C115	3107	45
C114	13	05/12/2009	86	179	C111-C112	148	4
					C111-C114	2838	43
					C111-C115	2789	57
C115	13	05/12/2009	86	179	C112-C114	1941	29
					C112-C115	1895	31
					C114-C115	281	4
C125	7	29/04/2010	92	253	C125-C126	736	27
					C125-C131	1159	22
C126	7	29/04/2010	92	253	C125-C133	872	22
					C125-C134	817	21
C131	7	08/05/2010	93	253	C126-C131	9636	99
					C126-C133	5847	71
C133	7	28/05/2010	94	253	C126-C134	5883	70
					C131-C133	3486	40
C134	7	19/12/2010	95	253	C131-C134	3400	38
					C133-C134	330	4
C156	16	16/11/2010	108	260	C156-C159	160	3
C159	16	09/12/2010	110	260			
C139	14	08/09/2010	98	ST2102	C139-C141	177	3
C141	14	01/10/2010	99	ST2102			

Supplementary Table S4 and Figure S1 are provided in additional files:

Table S4. Clone-specific deletions, relative to PAO1

Figure S1. Core genome SNP phylogeny showing the distribution of bronchiectasis isolates. The figure shows analysis of the genomes of all bronchiectasis isolates used in this study (highlighted in blue) alongside 331 genomes from Kos et al. [14] and the genomes of commonly studied strains PAO1 (labelled PAO1107), PA14 (UCBPPPA14109), PA7 and LESB58. Line colours indicate the two major clusters of *P. aeruginosa* (I, green; II, blue) as well as those isolates not clustering in the two main groups (red).

Figure S2. Distribution of loss of function mutations. For the genes listed in Table 2, where the number of independent occurrences of a loss of function mutation was equal to or greater than five, the Figure indicates the number of isolates carrying mutations in 0, 1, 2, 3, 4, 5 or 6 of these genes.

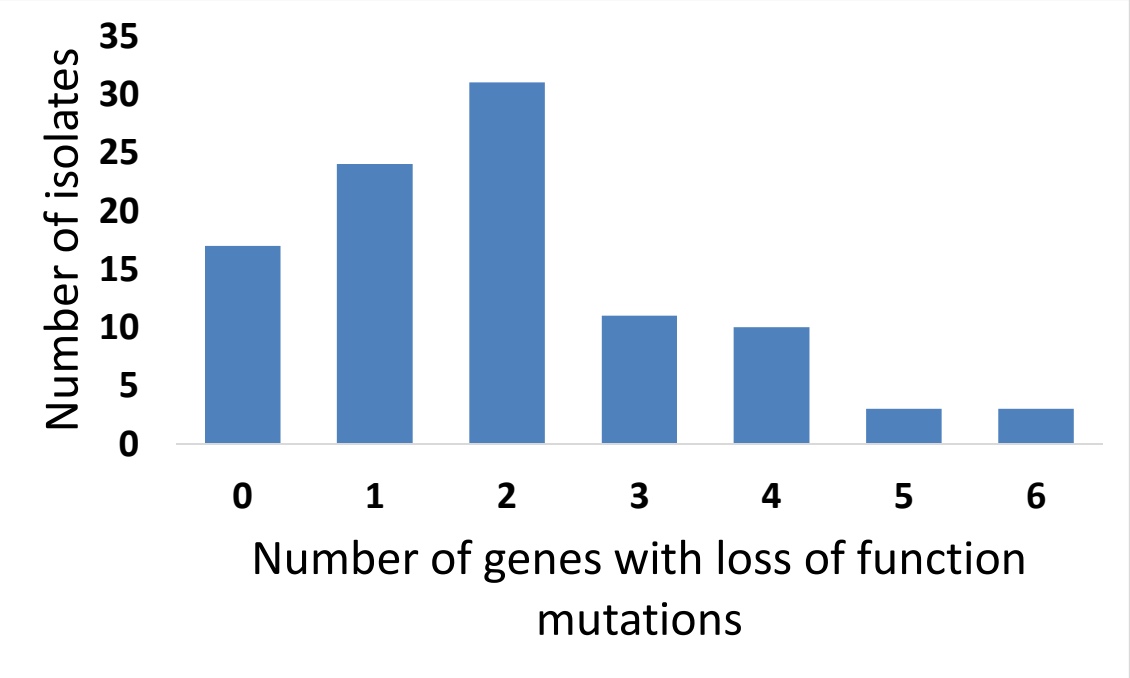


Table S5. Full list of loss of function mutations identified by variant calling. Available via the Figshare link https://figshare.com/s/ff426bae75ee64804aa1

Table S6. Loss of function mutations present in each bronchiectasis isolate genome. For the genes listed in Table 2, where the number of independent occurrences of a loss of function mutation was equal to or greater than five, the genes carrying such mutations are shown for each of the bronchiectasis isolates. 17 of the 99 isolates carried none of these mutations. It is worth noting that isolates found co-infecting individual patients did not share the same mutation profile (C125, C127

and C129 in patient 92; C12 and C13 in patient 42; C78 and C79 in patient 72; C86 and C87 in patient 73; C107 and C108 in patient 84; C109 and C110 in patient 85; A100 and A106 in patient 148).

Isolate	Number of Table2 genes with loss of function mutations	Mutations					
C125	6	PA4469	mexS	ladS	rbdA	betT2	mexA
C74	6	PA4469	mexS	chpA	mutL	pchE	mexB
C87	6	PA4469	bifA	pchE	ladS	chpA	mexA
C123	5	rbdA	PA4469	pchE	mexS	oprM	
C116	5	PA4469	mucA	mexB	oprF	pchE	
C88	5	mucA	pcoA	betT2	oprF	mexA	
C100	4	rbdA	PA4469	mucA	chpA		
C79	4	rbdA	PA4469	mucA	mutL		
C133	4	rbdA	mucA	pchE	oprM		
C95	4	rbdA	mucA	oprM	pcoA		
B16	4	rbdA	mucA	mexB	fimV		
C21	4	PA4469	mucA	mexB	pcoA		
C31	4	PA4469	mucA	mutL	fimV		
C110	4	PA4469	mucA	oprF	mexA		
C108	4	mexB	chpA	mucA	betT2		
C129	4	mucA	mexB	oprF	gmd		
C134	3	rbdA	mucA	oprM			
C143	3	rbdA	mucA	mexB			
C2	3	rbdA	mexB	chpA			
A119	3	PA4469	oprF	pcoA			
A46	3	PA4469	mucA	oprF			
C63	3	betT2	ladS	fimV			
C114	3	mexB	pchE	betT2			
A163	3	mucA	mexB	gmd			
C149	3	mucA	pchE	oprM			
C55	3	mucA	pilJ	oprF			
B113	3	pilJ	mutL	pcoA			
C124	2	PA0054	mucA				
C146	2	PA0054	mexA				
C49	2	PA0054	betT2				
C51	2	PA0054	rbdA				
B114	2	rbdA	mexB				
C25	2	rbdA	mexB				
C36	2	rbdA	mucA				
C86	2	rbdA	mexA				

C20	2	rbdA	pilJ
C137	2	rbdA	PA4469
C6	2	PA4469	mexS
C7	2	PA4469	mexS
C135	2	mucA	mexS
C61	2	PA4469	mexA
C42	2	PA4469	oprF
A100	2	PA4469	fimV
B62	2	PA4469	pilJ
C131	2	betT2	ladS
B34	2	bifA	betT2
C12	2	bifA	betT2
C11	2	bifA	betT2
A1	2	bifA	ladS
B3	2	fimV	pcoA
C167	2	mexA	gmd
C3	2	pilJ	mexB
A2	2	mucA	mexB
C153	2	mexB	oprF
C158	2	mexB	oprF
C161	2	mucA	pcoA
C18	2	oprM	mutL
C155	2	pilJ	oprM
C94	1	PA0054	
C120	1	rbdA	
A3	1	rbdA	
C77	1	rbdA	
C91	1	rbdA	
C69	1	rbdA	
C156	1	rbdA	
C159	1	rbdA	
C107	1	betT2	
C119	1	bifA	
A106	1	bifA	
C43	1	bifA	
B199	1	gmd	
C89	1	gmd	
C168	1	gmd	
A12	1	mexB	
C54	1	mexB	
C92	1	mucA	
C10	1	mucA	
C73	1	mucA	
C142	1	mucA	
C76	1	mucA	
B37	1	mucA	

C101 1 oprM

Table S7. The presence or absence of virulence genes amongst the non-CF BE isolate genomes. Available via the Figshare link https://figshare.com/s/626a59cfd94b13e5cf71

References

1. Martin M. Cutadapt removes adapter sequences from high-throughout sequencin reads. *EMBnet J* 2011: 17(1): 10-12.
2. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 2015: 31(4): 587-589.
3. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013: 29(8): 1072-1075.
4. Laing C, Buchanan C, Taboada EN, Zhang Y, Kropinski A, Villegas A, Thomas JE, Gannon VP. Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. *BMC bioinformatics* 2010: 11: 461.
5. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 2013: 30(12): 2725-2729.
6. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 2007: 23(1): 127-128.
7. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. *Genome biology* 2004: 5(2): R12.
8. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV. Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen. *Nature* 2000: 406(6799): 959-964.
9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009: 25(14): 1754-1760.
10. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009: 25(16): 2078-2079.
11. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 2010: 20(9): 1297-1303.
12. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 2012: 6(2): 80-92.
13. Mowat E, Paterson S, Fothergill JL, Wright EA, Ledson MJ, Walshaw MJ, Brockhurst MA, Winstanley C. Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections. *AmJRespirCrit Care Med* 2011: 183(12): 1674-1679.
14. Kos VN, Deraspe M, McLaughlin RE, Whiteaker JD, Roy PH, Alm RA, Corbeil J, Gardner H. The resistome of Pseudomonas aeruginosa in relationship to phenotypic susceptibility. *Antimicrobial agents and chemotherapy* 2015: 59(1): 427-436.